

Essays on Price Discovery

By

Cristina Mabel Scherrer

A Thesis submitted for the degree of
Doctor of Philosophy (Ph.D) in Economics

School of Economics and Finance

University of London

Queen Mary

June 2013

AUTHOR'S DECLARATION

I wish to declare:

No part of this doctoral dissertation, titled as “Essays on Price Discovery” and submitted to the University of London in pursuance of the degree of Doctor of Philosophy (Ph.D.) in Economics, has been presented to any University for any degree. Parts of Chapter 4 were undertaken as joint work with Professor Marcelo Fernandes.

Signed

Cristina Mabel Scherrer

ACKNOWLEDGEMENT

I thank my supervisor for his constant support, guidance and encouragement.

I thank my dear and lovely husband, without him this would not have been possible.

I thank my parents and my grandmother for teaching me the most important values in life.

I thank my sisters for being my best friends.

The only true wisdom is in knowing you know nothing. **Socrates**

EXTENDED ABSTRACT

Financial asset prices reflect investor’s perspectives over the current and future situation of a firm, an industry, a country and ultimately, the entire economy. For this reason, how financial asset prices are driven has been a fundamental economic question. Specific market characteristics such as the number of sellers and buyers, investors valuation perceptions, market availability of other assets and legal and technical properties are some of the features that affect asset prices. When the same asset is traded at different venues, these specific characteristics may vary, following a certain degree of heterogeneity across buyers and sellers. The direct consequence is that transaction prices of the same asset differ across markets. However, prices will also not drift apart, since arbitrage opportunities would arise, reducing or even eliminating the differences. Prices of similar securities linked to a single latent price, as derivative markets, for instance, present the same behaviour. Price differences among markets observed at high frequencies are an indication that venues incorporate new information in an unlike way. The structure and design of a market impacts its behaviour, liquidity, efficiency, and hence how prices are discovered. The task of identifying the leading markets and understanding how the price dynamics occurs are the main objectives of the price discovery analysis.

Chapter 1 introduces the research subject of price discovery, motivating the importance of what this thesis proposes and the results and conclusions obtained.

Chapter 2 explains in details the main methodologies used to measure price discovery and the important results in the empirical literature.

Chapter 3 motivates the data set this thesis uses, with institutional

background details and specific market and firm characteristics. We also present in details the steps we follow to deal with standard issues of high frequency data, such as outliers and errors on a tick-by-tick database and non synchronicity of prices at different markets.

Chapter 4 extends the standard price discovery model to estimate the information share (IS) accounting for the information content of both common and preferred non US stocks, their American Depositary Receipts (ADRs) counterparts traded on the New York Stock Exchange and ARCA, and the exchange rate. We gauge the significance of price discovery in the home and foreign markets, through common or preferred stocks. One of the main critiques on the IS methodology is that it does not deliver a single measure when there is contemporaneous correlation among markets. We propose an ordering invariant methodology that delivers a single measure of IS. We find that the foreign market is more important than the home market for the price discovery of Petrobras, the Brazilian stated-owned oil giant, and Vale, one of the largest mining companies in the world. Additionally, the Brazilian market has lost significant importance after the 2008/2009 financial crisis. During this period, common and preferred stocks shared a single common factor, with voting premium being a stationary process.

Chapter 5 investigates instantaneous and long-run linkages between common and preferred shares traded at both domestic and foreign markets. We develop a market microstructure model in which the dynamics of the different share prices react to three common factors, namely, the efficient price, the efficient exchange rate, and the efficient voting premium. We show how to identify the structural innovations so as to differentiate instantaneous and long- run effects. First, we obtain dynamic measures of

price discovery that quantify how prices traded at different venues respond to shocks on the common factors. Second, we are able to test whether shocks in the efficient exchange rate change the value of the firm. Third, we test whether shocks on the efficient voting premium have a permanent effect on preferred shares. We implement an empirical application using high-frequency data on six Brazilian large companies. We find that, in the long-run, a depreciation of the Brazilian currency leads to a depreciation of the value of the firm that exceeds the expected arbitrage adjustment. In addition, a positive shock on the voting premium yields a positive impact on the value of the firm. Our price discovery analysis also reveals that one trading day suffices to impound new information on all share prices, regardless of the venue they trade at.

Finally, Chapter [6](#) concludes.

Contents

1	Introduction	11
2	The incorporation of news: Price Discovery	17
2.1	Efficient Market and the Random Walk Model	19
2.2	Measures of Price Discovery	25
2.2.1	Permanent Temporary (PT) Decomposition	26
2.2.2	Information Share (IS)	37
2.2.3	Comparison: Permanent Temporary (PT) and Infor- mation Share (IS)	45
2.3	Empirical Literature Review	61
2.3.1	Low Frequency Data	62
2.3.2	High Frequency Data	71
3	Dealing with a High Frequency Data Set	80
3.1	Institutional Background	80
3.2	Data Description	83
3.3	Data cleaning and aggregation	85
4	An order invariant Measure for Price Discovery	102
4.1	Introduction	102

4.2	Information share in a large price system	107
4.3	Which share class leads, and in which market?	111
4.3.1	Petrobras	113
4.3.2	Vale	117
4.4	Simulations	119
4.5	Conclusion	122
5	A Dynamic Measure for Price Discovery	132
5.1	Introduction	132
5.2	A simple model for price discovery	139
5.3	Getting structural parameters from reduced-form VECM	148
5.4	Price discovery for Brazilian cross listed stocks	157
5.5	Robustness	162
5.5.1	Rolling Window	163
5.5.2	Interval Frequency	164
5.6	Identification Issues	164
5.7	Simulations	171
5.8	Conclusion	174
6	Conclusion	191

List of Figures

3.1	Intersection in trading hours (São Paulo time, UTC −3 hours)	91
3.2	Price Evolution	99
3.3	The prices of Petrobras' shares and their ADR coun- terparts	100
3.4	The prices of Vale's shares and their ADR counter- parts	101
5.1	Bradesco	185
5.2	Gerdau	186
5.3	Ambev	187
5.4	BR Telecom	188
5.5	Vale	189
5.6	Petrobras	190

List of Tables

3.1	Data set size and number of entries	90
3.2	2008 Most traded stocks at Bovespa	92
3.3	2009 Most traded stocks at Bovespa	93
3.4	Liquidity of Bovespa across NYSE opening times . .	94
3.5	Sample sizes before and after discarding outlier . . .	95
3.6	Data Cleaning Details	96
3.7	Aggregating tick data into fixed time intervals	97
3.8	Data Aggregation	98
4.1	Information share for Petrobras in 2008	125
4.2	Information share for Petrobras in 2009	126
4.3	Information share for Vale	127
4.4	Information share for M_1 with a diagonal covariance matrix	128
4.5	Information share for M_1 with a nondiagonal covari- ance matrix	129
4.6	Information share for M_2 with a nondiagonal covari- ance matrix	130
4.7	Information share for M_3 with a nondiagonal covari- ance matrix	131

5.1	Price Discovery BR Telecom 300”	176
5.2	Price Discovery BR Telecom 240”	177
5.3	Price Discovery Bradesco 30”	178
5.4	Price Discovery Gerdau 30”	179
5.5	Price Discovery Gerdau 60”	180
5.6	Price Discovery Ambev 90”	181
5.7	Price Discovery Petrobras 30”	182
5.8	Price Discovery Vale 30”	183
5.9	Monte Carlo Simulations	184

Chapter 1

Introduction

One of the functions of a market is to determine the “correct” price of an asset. The interaction between buyers and sellers determine the market price. Specific market characteristics such as the number of sellers and buyers, investors valuation perceptions, market availability of other assets and legal and technical properties are some of the features that affect asset prices. When the same asset is traded at different venues, these specific characteristics may vary, following a certain degree of heterogeneity across buyers and sellers. The direct consequence is that transaction prices of the same asset differ across markets. However, these prices do not drift apart, since arbitrage considerations should eliminate price differences quickly enough. Furthermore, there still exist a non-arbitrage band in the short-run, where a given market incorporates new information first, relative to others. The availability of ultra high frequency data allows a much more accurate analysis of the intraday price co-movement. The main reason for this gain arises because daily intervals may be too long, implying that equilibria among different markets are restored by the time markets close. The speed

at which markets reach a new equilibrium may also change with massive improvements on information technology, financial markets liberalization and increase in the use of algo-trading strategies, as we have experienced in the past years. Considering such environment, the task of identifying the leading market, quantifying its importance, finding the short-run price dynamics, and ultimately, the price discovery mechanism, have been research topics of central importance.

This thesis investigates how assets driven by the same fundamentals and traded at various markets react to innovations. We work with a high-frequency data on large Brazilian companies traded at BM&FBovespa (the Brazilian stock exchange), New York Stock Exchange (NYSE) and Archipelago Exchange (ARCA). These firms are traded at the US markets as American Depositary Receipts (ADR). This data set is particularly suitable for this study for two reasons. First, there is a large trade intersection period in the entire year between Brazil and US exchanges, varying from six hours and thirty minutes to five hours and thirty minutes. This yields a larger amount of information regarding the price discovery process, when compared to studies among companies traded at the European and the US markets. In these studies, the intersection period is merely from two to three hours. Second, we work with companies that possess distinctive characteristics such as: core business, ownership structure, global insertion, and strategic and political relevance, providing a wider picture to check whether the model proposed delivers plausible conclusions. This allows wider conclusions and not industry or sector specific results.

In Chapter 2 we present a literature review, which encompasses a detailed explanation of the two main methodologies to measure price dis-

covery. The importance of this review comes because this is a starting point for the development of two different approaches covered in Chapters 4 and 5. We also present a brief review on some major empirical results this literature has provided. In terms of tools to measure price discovery, the literature has been lacking of an order invariant methodology for a static analysis of price discovery in reduced form models, as well as for a dynamic measure in structural models. Methodologically, these are the two main contributions of the present thesis (Chapter 4 and 5 present the proposed measures).

Chapter 3 presents in detail the data set used in this thesis and the institutional background surrounding it. We explain the necessary steps to deal with a high frequency data set. Basically, the two important steps are to clean the data and to deal with non synchronicity. Raw ultra high frequency data may present some behaviour which is not consistent with standard market activity. Given that, we implement a methodology to withdrawn outliers of the data set. Secondly, we aggregate the time series at different time stamps. This chapter presents the details of the methodologies used, as well as, the results found.

Chapter 4 extends the standard price discovery analysis to estimate the information share of dual-class shares across domestic and foreign markets. By examining both common and preferred shares, we aim to extract information not only about the fundamental value of the firm, but also about the dual-class premium. In particular, our interest lies on the price discovery mechanism regulating the prices of common and preferred shares in the BM&FBovespa as well as the prices of their ADR counterparts in the NYSE and in the Arca platform. However, in the presence of contem-

poraneous correlation between the innovations, the standard information share measure depends heavily on the ordering we attribute to prices in the system. To remain agnostic about which are the leading share class and market, one could for instance compute some weighted average information share across all possible orderings. This is extremely inconvenient given that we are dealing with 2 share prices in Brazil, 4 share prices in the US, plus the exchange rate (and hence over 5,000 permutations!). We thus develop a novel methodology to carry out price discovery analyses that does not impose any ex-ante assumption about which share class or trading platform conveys more information about shocks in the fundamental price. As such, our procedure yields a single measure of information share, which is invariant to the ordering of the variables in the system. Simulations of a simple market microstructure model show that our information share estimator works pretty well in practice. We then employ transactions data to study price discovery in two dual-class Brazilian stocks and their ADRs. We uncover two interesting findings. First, the foreign market is at least as informative as the home market. Second, shocks in the dual-class premium entail a permanent effect in normal times, but transitory in periods of financial distress. We argue that the latter is consistent with the expropriation of preferred shareholders as a class.

Chapter 5 develops a new dynamic measure for price discovery. We show that price's fundamentals may have cross linkages, meaning that an innovation on exchange rate may have a permanent effect on the latent asset efficient price, for instance. We document a novel conclusion in the dynamic price discovery analyses: an innovation of one unit on a price fundamental may have an impact larger than that on observed prices, given cross linkages

among fundamentals. This is a major contribution to the price discovery literature, because no study has considered cross linkages among common factors before, rendering potential misleading conclusions.

We investigate instantaneous and long-run impacts on common and preferred shares traded at both domestic and foreign markets. We set up a novel price discovery model allowing for cross linkages among three common factors, namely, the efficient price, the efficient exchange rate, and the efficient voting premium. In this model, we isolate instantaneous and long-run effects on prices given structural innovations on the common factors. Our flexible econometric specification allows us to look at cross linkages in the same way the theoretical model does. We show how to identify the structural innovations with minimal restrictions. Our theoretical model and econometric methodology yield three developments: First, we obtain dynamic measures of price discovery that quantify how prices traded at different venues respond to shocks on the common factors over time. Second, we are able to test whether shocks in the efficient exchange rate change the value of the firm exceeding the expected arbitrage adjustment. Third, we test whether shocks on the efficient voting premium have a permanent effect on preferred shares. The latter two are assessed without imposing any firm-specific assumptions, such as production function, market orientation, financing and ownership structures.

The empirical results in this chapter corroborate the solutions the price discovery model proposed, implying that innovations associated with the latent processes are contemporaneously correlated, leading to cross linkages among common factors. This shows that measuring price discovery independently of exchange rate or other common factors may lead to misleading

results. We document that, in the long-run, a depreciation of the Brazilian currency leads to a depreciation of the value of the firm that exceeds the expected arbitrage adjustment. In addition, a positive shock on the voting premium yields a positive impact on the value of the firm. In general, ARCA is faster than NYSE in the short run, but they are equally important in the long run. These results are consistent across all the six companies, as well as at different sampling frequencies. Our price discovery analysis also reveals that one trading day suffices to impound new information on all share prices, regardless of the venue they trade at.

Finally, Chapter 6 concludes.

Chapter 2

The incorporation of news: Price Discovery

Studying how financial asset prices are affected by news, and ultimately how they are formed is a key factor in financial market analysis. Understanding how innovations affect financial assets prices, as well as when they do, comprises a major interest of financial economists.

The scenario where a same or similar asset is traded at different venues is quite common. Derived from that, it comes to interest how these prices link to each other and how they relate to news in general. Following the concept of efficient market theory and that prices follow a random walk process (or at least a component of prices - the unobservable efficient price - is modelled as a random walk) and therefore only new information can affect prices, our interest moves to how news are incorporated on prices. Thus, the questions are: when, where and how the inclusion of news happens on the existing price, driving it to a price change. Comprehension and inference of which market is responsible for first incorporating news to its price is a subject

titled as price discovery.

Hasbrouck (1995) points out the importance of understanding where the price discovery (which he defines as “the incorporation of new information” or as “the impounding of new information into the security price”) occurs when a given security is traded at different venues. One may extend this analysis for the case of a security and its derived securities (such as spot and futures market, stocks and options). In these cases, the security prices present the same efficient price, that is, securities share a common factor, also defined as a common unobserved efficient price.

Lehmann (2002) notices the importance of secondary markets to price discovery:

“One of the central functions of secondary markets is price discovery: the efficient and timely incorporation of the information implicit in investor trading into market prices.” Lehmann (2002), page 259

Garbade and Silber (1979) explain the role of costs on price divergence between different markets. Considering extreme high transportation cost and making communication close to impossible, prices are unrelated and thus markets are independent. On the other hand, however, when costs are null and communication is widely spread, markets are perfectly integrated and prices are identical. The in-between case is when markets are imperfectly integrated. In such a case, the fact which determines how fast price adjustments occur is communication technologies and institutional arrangements. There are two ways this price adjustment can occur: symmetrically (prices in market A change towards prices in market B as fast as prices in market B change towards the ones in market A) or in one way (prices in

market B adjust to prices in market A with some delay, which is defined by communication speed. They define the later case as a dominant-satellite market relationship (market A being the dominant market, and market B being the satellite). They perform some empirical work with USA data from NYSE and two regional stock exchanges (August, 1973 - September 1975), and find a suggestion to reject the null hypothesis of these exchanges being perfectly integrated. They also observe an evidence that the regional exchanges are satellites of the NYSE.

There are some important steps to follow in order to get the full understanding of price discovery and its significance. Beginning this discussion with the market efficiency theory and the standard random walk model is a must. From this starting point, we continue the analysis with the theoretical framework and econometrics tools of how one should measure market's importance on price determination. We make this analysis in details for the two most used measures of price discovery, namely the component share (from the permanent temporary decomposition) and the information share. We then go to the empirical application and interesting results found in the literature using high and low frequency data.

2.1 Efficient Market and the Random Walk Model

Nowadays, although it is known that the basic random walk model is not always a complete way to describe and to understand stocks prices (for instance, when the observation intervals are short), it is a very good starting point and at least part of the price formation comprehension. Dealing with

high frequency data may bring some microstructure effects, that arise due to imperfections on the trade process, such as discreteness of prices, properties of the trading mechanisms, informational effects, bid ask bounces, inventory dealing effects and others. However, even in the above cases, random walk models are still very useful as a way to define the unobserved efficient price, instead of the observed market price, as we shall see later. The complete picture of random walk processes as a framework to model securities' prices comes together with the market efficiency theory and the work of Fama (1965).

Two very important topics are worth highlighting on the work of Fama (1965). The first one relates to the possibility of having a very good estimate of the intrinsic value of a security, under the efficient market theory. The second one goes further by explicating that prices reflect all current information available and they adjust fast to new information. Therefore, we can assert that only new information is capable to change current prices.

A large number of buyers and sellers seeking profit maximization in a competitive market form a basis for an efficient market. Fundamental analyses find a security's intrinsic value by evaluating specific characteristics of this asset, such as past decisions made by the company, future expected cash flows, business environment, etc. At an efficient market, investors look and analyze all available information, and after an interaction of traders with heterogeneous beliefs and preferences, a consensus on the market security's value arises as the market price. Following that, the market price would be a fair estimate of the intrinsic value of this security. Since investors look at all available information (past and projections), prices reflect not only historical events, but future events, which is made

of investors estimates.

Expanding on this concept, a market is considered efficient when information is freely available to everyone, there are no trading costs and investors are aware that information knowledge affects prices. In this way, current security's price incorporates all available information at the current moment. As a consequence, only new information could affect the current price. New information can be actual or anticipated information, delivering some uncertainty to the new intrinsic value. This may cause, together with the instantaneous adjustment (since this is a competitive market), some over or under adjustment. The lag in the final adjustment is itself a random and independent variable. As the information can be anticipated by estimates, the actual price sometimes may adjust prior to the future event, and other times, later. The property of instantaneous adjustment implies that new information incorporated to prices (or in other words, security price changes) is independent. Thence, such a process is a random walk model¹, as stated in (2.1).

$$P_t = P_{t-1} + \varepsilon_t \tag{2.1}$$

Where ε_t is an independent and identically distributed (*i.i.d*) variable, with zero mean and variance equal to 1. Prices following a random walk process possess returns which are unpredictable and therefore independent over time. Hence, under the efficient market theory, the best prediction for

¹A random walk process is a special case of a martingale process, which can be defined as $E(X_{t+s} | I_t) = X_t \iff E(X_{t+s} - X_t | I_t) = 0 \ (\forall s > 0)$ and $(E | X_t | < \infty, \forall t)$. Prices following a martingale model were considered a necessary condition for an efficient market in the past. However, some issues raised (such as the fact that this model would not allow for a risk return analysis) made the description of price process to be extended to a random walk model, since the latter one allows the inclusion of a drift to explain normal profits.

tomorrow's price, is today's price, since the expected value of ε_t term is equal do zero.

The random walk definition above is one of the three possible ways to define this type of process. Instead of making the assumption of *i.i.d* increments (ε_t is an *i.i.d* process), this assumption could be replaced by ε_t being independent but not identically distributed, or even that ε_t is only an uncorrelated (weak white noise) process, but presenting other forms of dependency. These two other assumptions are weaker assumptions on the increments of random walk models.

Hasbrouck (2002) points out the importance of the random walk model, even when one is dealing with short intervals on stock returns. In such cases, there is the presence of microstructure effects, which includes bid ask bounces, discreteness of price changes, etc. The random walk model may not be sufficient to describe prices movements when the influence of these effects is large. However, even then, there is an implicit random walk component that although does not describe prices movements in its full, still has economic importance. Hence, instead of considering trading prices following a random walk, one supposes the efficient price (or the intrinsic value) to follow a random walk process, as below:

$$m_t = m_{t-1} + v_t \tag{2.2}$$

Where v_t is an *i.i.d* variable, with zero mean and variance equal to 1. m_t is not observable. The observed price (the transaction price) is then equal to the unobserved term plus microstructure effects.

$$p_t = m_t + s_t \tag{2.3}$$

Fama (1970) defines an efficient market deeply, as an instance where prices always fully reflect all available information. He points out the possible causes of market inefficiency highlighting their consequences in real world:

“But though transactions costs, information that is not freely available to all investors, and disagreement among investors about implications of given information are not necessarily sources of market inefficiency, they are potential sources. And all three exist to some extent in real world markets. Measuring their effects on the process of price formation is, of course, the major goal of empirical work in this area.” Fama (1970), page 388

Fama (1970) also brings some review on the efficient market theory, creating a link between the theory and empirical work done at that time. He separates the empirical work on testing market efficiency in three groups: weak, semi-strong and strong form tests. The weak form tests consider the information set as being only the past history of prices, the semi strong takes into consideration a wider information set, as it includes all information known to all market participants, such as financial statements and economic conditions. Finally, the strong form tests also include private information besides public information. He finds no relevant evidence against the efficient market theory when weak and semi-strong form tests are used, and finds little evidence, when the strong form test is considered.

Roll (1984) uses the efficient market preposition and inserts cost in the trading process, developing a measure of bid ask spread.

“When transactions are costly to effectuate, a market maker

(or dealer) must be compensated; the usual compensation arrangement includes a bid-ask spread, a small region of price which brackets the underlying value of the asset. The market is still informational efficient if the underlying value fluctuates randomly. We might think of “value” as being the center of the spread. When news arrives, both bid and ask prices move to different levels such that their average is the new equilibrium value. Thus, the bid-ask average fluctuates randomly in an efficient market.” Roll (1984), page 1128

He considers a random walk model for the unobserved efficient price as in 2.2 and a market conducted by dealers, with a cost per trade equal to c (a constant term). As a result, trade price is equal to the efficient price plus the trade cost c (in the case of a ask) or minus the trade cost (for a bid). This makes the bid-ask spread equals to (2.4).

$$\text{bid-ask} = m_t + c - (m_t - c) = 2c \quad (2.4)$$

The transaction price can be written as below:

$$p_t = m_t + q_t c \quad (2.5)$$

Where q_t is equal to +1 (if investor in buying, ask quote) and q_t is equal to -1 (if investor is selling, bid quote). With the above assumptions, and assuming that buys and sells are equally likely, serially independent and not related to price changes in the efficient price (v_t), the so called Roll Model finds an expression to the bid ask spread by calculating trade price changes (Δp_t) variance and covariance. Finally, Roll (1984) finds that

$cov(\Delta p_t) = -c^2$, hence:

$$Spread = 2\sqrt{-cov(\Delta p_t)} \quad (2.6)$$

Some empirical applications compare results to firm size, being strongly negatively related to it. He finds some evidence of informational inefficiency, with different spread estimates when dealing with weekly and daily data.

Hasbrouck (2007) points out some well known questions to the previous assumptions. Using empirical data he shows that serial dependency between q_t and q_{t-1} can occur in financial data (buys tend to follow buys, and the same for sells) as well as the existence of dependency between v_t (in (2.2)) and q_t (in (2.5)), since changes in an asset's intrinsic value can interfere in trade direction.

2.2 Measures of Price Discovery

The interest in price discovery led to a development of different methodologies to try to measure the importance of different markets in an asset's price formation. Empirical studies have followed two main econometric methodologies: the Permanent Temporary Decomposition from Gonzalo and Granger (1995) (which was applied to the price discovery concept by Booth, So, and Tseh (1999), Chu, Hsieh, and Tse (1999) and Harris, McInish, and Wood (2002), measuring the component share (CS) of each market in the price formation) and the Information Share (IS) from Hasbrouck (1995). We analyze both of them in a deeper way in the following two subsections. On the third subsection, we go over some studies comparing

the two methodologies, almost all from the special issue of the Journal of Financial Markets 5 in 2002.

2.2.1 Permanent Temporary (PT) Decomposition

In order to explain Gonzalo and Granger (1995) approach, it is necessary firstly to go over the cointegration definition and concept. In what follows, we detail the permanent temporary decomposition using methodology and explanations from Gonzalo and Granger (1995).

Two non stationary ² variables and integrated of order one are called cointegrated time series if there is a linear combination of these two variables which is a stationary process. For instance, consider a vector of variables y_t , where all variables are integrated of order 1, or I(1). If one is able to find a linear combination of these variables presenting a characteristic of stationarity, that is, integrated of order zero or I(0), such as βy_t , then, y_t is a cointegrated process and β is a cointegrating vector.

If any two series are cointegrated, there is a common factor representa-

²A weakly stationary process (y_t) presents its first and second moments as time invariant. This means:

$$E(y_t) = v$$

for all t ; and

$$E[(y_t - v)(y_{t-h} - v)'] = \Gamma_y(h) = \Gamma_y(-h)'$$

for all t and $h = 0, 1, 2, \dots$

A strictly stationary process (y_t) has the joint distribution of (y_{t1}, \dots, y_{tk}) equal to the joint distribution of $(y_{t1+t}, \dots, y_{tk+t})$ for all t (where t is a positive integer). As this is a very strong assumption, we work with the weakly stationarity definition.

tion between these two series (from Stock and Watson (1988)), as follow:

$$\begin{bmatrix} y_t \\ x_t \end{bmatrix} = \begin{bmatrix} A \\ 1 \end{bmatrix} f_t + \begin{bmatrix} \tilde{y}_t \\ \tilde{x}_t \end{bmatrix} \quad (2.7)$$

The common factor f_t is integrated of order 1, and \tilde{y}_t and \tilde{x}_t are stationary. Gonzalo and Granger (1995) point out that many research has been done in order to estimate the cointegrated vector $(1, -A)$ ³ as a way to understand the long run relation between the two variables, however not much attention has been given to the common factor estimation, f_t . They bring the attention to the importance of estimating f_t and not just the cointegrating vector. By estimating f_t , one may reduce the number of estimated parameters, by reducing the number of variables, which can be very useful in a large cointegration system. Another reason is that one is able to split the system in two different components: the permanent (f_t), and the transitory ones $(\tilde{y}_t, \tilde{x}_t)$.

For these components to be identified, one needs to impose some conditions: f_t to be a linear combination of (y_t, x_t) (which makes f_t observable) and $(\tilde{y}_t, \tilde{x}_t)$ not to have any permanent component. That is, every permanent effect on (y_t, x_t) is restricted to the common factor expression, which assures f_t is a good option to express all the long term behavior.

Before we continue with the factor model from Gonzalo and Granger (1995), it is important to go over the Error Correction Model (ECM). The

³Considering long term equilibrium, one could express the vector $\begin{bmatrix} y_t \\ x_t \end{bmatrix}$ only with the I(1) common factor term, as below
 $\begin{bmatrix} y_t \\ x_t \end{bmatrix} = \begin{bmatrix} A \\ 1 \end{bmatrix} f_t$, $y_t = A f_t$ and $x_t = f_t$. Hence, $y_t - A x_t = (1 - A) \begin{pmatrix} y_t \\ x_t \end{pmatrix}$.

explanation below follows Lutkepohl (2007).

An error correction model describes a long run equilibrium where variations occur around this equilibrium and depend on it. As an example, consider two different venues trading the same asset, where y_{1t} is the asset's price in the first market and y_{2t} in the second. There is an equilibrium relation between these two prices, expressed by $y_{1t} = \beta_1 y_{2t}$. The equation below shows how variations in prices in the first market (from period $t-1$ to period t , for instance) can be written as a function of deviations from the equilibrium relation in period $t-1$ plus a white noise term.

$$\Delta y_{1t} = \alpha_1 (y_{1,t-1} - \beta_1 y_{2,t-1}) + u_{1t} \quad (2.8)$$

The price change in the second venue may also be expressed as a function of the equilibrium deviation plus a white noise term.

$$\Delta y_{2t} = \alpha_2 (y_{1,t-1} - \beta_1 y_{2,t-1}) + u_{2t} \quad (2.9)$$

One may extend the two equations above to a more general form, where Δy_{1t} depends also on other terms.

$$\Delta y_{1t} = \alpha_1 (y_{1,t-1} - \beta_1 y_{2,t-1}) + \gamma_{11,1} \Delta y_{1,t-1} + \gamma_{12,1} \Delta y_{2,t-1} + u_{1t} \quad (2.10)$$

$$\Delta y_{2t} = \alpha_2 (y_{1,t-1} - \beta_1 y_{2,t-1}) + \gamma_{21,1} \Delta y_{1,t-1} + \gamma_{22,1} \Delta y_{2,t-1} + u_{2t} \quad (2.11)$$

Considering y_{1t} and y_{2t} integrated of order one variables, the first difference of them (Δy_{1t} and Δy_{2t}) is stationary, that is, $I(0)$ variables. If Δy_{1t} and Δy_{2t} are stationary and stable variables, all Δy_{it} terms in (2.10) and in (2.11) must also be stationary, as well as the white noise terms, u_{1t} and

u_{2t} . Hence, moving all stable terms (all Δy_{it} terms and white noise terms) to one side of the equation, we end up with the following:

$$\alpha_i (y_{1,t-1} - \beta_1 y_{2,t-1}) = \Delta y_{it} - \gamma_{i1,1} \Delta y_{1,t-1} - \gamma_{i2,1} \Delta y_{2,t-1} - u_{it} \quad (2.12)$$

There is a group of stable terms in the right side of the equation, which equals to the left side of the equation that also needs to be stable, as there is no possibility of a stable term being equal to an unstable term. Therefore, considering $\alpha_i \neq 0$, $y_{1,t-1} - \beta_1 y_{2,t-1}$ must be stable, and consequently constitute a cointegration term. Also, (2.10) and (2.11) may be written in a matrix and vector notation⁴.

$$y_t - y_{t-1} = \alpha \beta' y_{t-1} + \Gamma_1 (y_{t-1} - y_{t-2}) + u_t \quad (2.13)$$

Or in a more simplified way

$$\Delta y_t = \alpha \beta' y_{t-1} + \Gamma_1 \Delta y_{t-1} + u_t$$

where $y_t = (y_{1t}, y_{2t})'$, $u_t = (u_{1t}, u_{2t})'$, $\alpha = [\alpha_1, \alpha_2]'$, $\beta' = (1, -\beta_1)$, and $\Gamma_1 = \begin{bmatrix} \gamma_{11,1} & \gamma_{12,1} \\ \gamma_{21,1} & \gamma_{22,1} \end{bmatrix}$.

Now that we have defined an error correction model, we may go back to Gonzalo and Granger (1995).

First, assuming X_t is a vector ($p \times 1$) of integrated of order 1 series with

⁴Equation (2.13) could also be written using a VAR(2) representation, just by rearranging terms, as below:

$$y_t = (I_k + \Gamma_1 + \alpha \beta') y_{t-1} - \Gamma_1 y_{t-2} + u_t$$

mean equal to zero⁵, and there is a matrix $\alpha_{p \times r}$ of rank r which makes $\alpha' X_t$ to be $I(0)$. Hence, X_t variables are cointegrated and can be represented by a error correction model:

$$\Delta X_t = \gamma \alpha' X_{t-1} + \sum_{i=1}^{\infty} \Gamma_i \Delta X_{t-i} + \epsilon_t \quad (2.14)$$

where γ is a $(p \times r)$ matrix and α is a $(r \times p)$ matrix.

Following (2.7), X_t can be written as the sum of a permanent term, $I(1)$ and a transitory term, $I(0)$, as stated in (2.15) below.

$$X_t = A_1 f_t + \widetilde{X}_t \quad (2.15)$$

where X_t is a $(p \times 1)$ matrix, A_1 is $(p \times k)$, f_t is $(k \times 1)$ and \widetilde{X}_t is $(p \times 1)$.

The number of common factors is given by the dimension of f_t matrix, where $k = p - r$. As $A_1 f_t$ is an $I(1)$ expression, and there are r linear combinations that make X_t an $I(0)$ variable, the dimension of f_t is the total number of possible common factors, p (which is the dimension of X_t) minus the number of linear combinations that make X_t an $I(0)$ variable. Hence, we have $k = p - r$. Intuitively, in one hand, $A_1 f_t$ is able to express all the $I(1)$ components of X_t (long run feature). On the other hand, the r linear combinations express all short run feature (which is adjusted by γ in order to match dimensions) resulting in the \widetilde{X}_t expression, with dimension $(p \times 1)$.

Gonzalo and Granger (1995) say that in the standard factor analysis, the interest usually is on estimating the matrix A_1 and the number of

⁵Gonzalo and Granger (1995) use this assumption in order to simplify the analysis. However, if one works with price series, it would not make sense to assume zero mean for the price series.

common factors k . These are already known from (2.14) estimation and from that fact that $\alpha' A_1 = 0$. The main purpose of Gonzalo and Granger (1995) is to estimate the common factor f_t using equation (2.14) instead of using (2.15), as it is more often done in factor analysis.

As said before, the two conditions added to identify the common factors are:

$$f_t = B_1 X_t \quad (2.16)$$

where f_t is $(k \times 1)$, B_1 is $(k \times p)$, X_t is $(p \times 1)$; and that $A_1 f_t$ composes the permanent component of X_t and \widetilde{X}_t composes the transitory effect.

Below, Definition 1 is stated as in their work.

Consider X_t an $I(1)$ series. A Permanent - Transitory decomposition for X_t is composed by two stochastic process (P_t and T_t), such that:

1. P_t is a difference stationary ($I(1)$) and T_t is a covariance stationary ($I(0)$);
2. $var(\Delta P_t) > 0$ and $var(T_t) > 0$;
3. $X_t = P_t + T_t$
4. Autoregressive representation (AR) of ΔP_t and T_t , considering u_{pt} and u_{Tt} uncorrelated.

$$H^*_{(p \times p)}(L) \begin{bmatrix} \Delta P_t \\ T_t \end{bmatrix} = \begin{bmatrix} u_{Pt} \\ u_{Tt} \end{bmatrix} \quad (2.17)$$

Also,

$$\lim_{h \rightarrow \infty} \frac{\partial E_t(X_{t+h})}{\partial u_{P_t}} \neq 0 \quad (2.18)$$

$$\lim_{h \rightarrow \infty} \frac{\partial E_t(X_{t+h})}{\partial u_{T_t}} = 0 \quad (2.19)$$

with E_t being the conditional expectation with relation to past information.

The last condition says that innovations from the permanent component u_{P_t} affects the long run forecast of the variable X_t , while impacts from a transitory effect (u_{T_t} , as in the second limit) do not affect variable X_t forecast.

In others words, the last condition can be stated as below:

$$\begin{bmatrix} H_{11}(L) & H_{12}(L) \\ H_{21}(L) & H_{22}(L) \end{bmatrix} \begin{bmatrix} \Delta P_t \\ T_t \end{bmatrix} = \begin{bmatrix} u_{1t} \\ u_{2t} \end{bmatrix} \quad (2.20)$$

It is necessary to assume that the total multiplier of ΔP_t with respect to the transitory component has to be equal to zero ($H_{12}(L) = 0$). This means that if we consider a series split into permanent and transitory effects, we have the following equations:

$$X_t = P_t + T_t \quad (2.21)$$

$$\Delta P_t = a_1 T_{t-1} + a_2 \Delta T_{t-1} + u_{1t} \quad (2.22)$$

$$T_t = b_1 \Delta P_{t-1} + u_{2t} \quad (2.23)$$

As a way to have a permanent transitory decomposition, a_1 must be equal to zero, in order to the transitory component not cause any permanent effect on X_t . If this is not true, the transitory component leads to a

permanent effect on X_t . By analyzing these equations, one can see that changes in the transitory component may have an impact on changes in the permanent component, as well as, changes in the permanent component can affect the transitory term.

On their work, some propositions are also presented, in order to have the model better specified. In summary, these propositions show that the conditions stated before (f_t is a linear combination of (y_t, x_t) and $(\tilde{y}_t, \tilde{x}_t)$ not have any permanent component) are sufficient to identify the common factors, common factor decomposition existence, and finally, the last proposition assures that the random walk component of the I(1) common factor f_t in the equation below

$$X_t = A_1 f_t + A_2 z_t$$

corresponds to the common trend of the Stock Watson decomposition.⁶

The common-trends representation follows:

$$X_t = X_0 + A\tau_t + a_t$$

The equation above comes from the cointegrated vector moving average representation as:

$$\Delta X_t = \mu + C(L)\varepsilon_t$$

⁶According to Stock and Watson (1988):

“The common-trends representation expresses X_t as a linear combination of k random walks with drift π , plus some transitory components a_t that are integrated of order 0.” page 1098

One assumes that the matrix $C(1)$ has rank equal to $k < n$, so X_t is cointegrated. If $k = n$, the variables would not be cointegrated, as α , which its columns are the cointegrated vector of X_t , has a rank equal to $r = n - k$. If $k = n$, r would be equal to zero, then no such a matrix α would exist. Matrix α is such that $\alpha' C(1) = 0$ and $\alpha' \mu = 0$. A representation for X_t from the equation above is then:

$$X_t = X_0 + \mu_t + C(1) \sum_{s=1}^t \varepsilon_s + C^*(L) \varepsilon_t$$

One can find the stationary linear combinations by multiplying the cointegrating vectors α' .

$$Z_t = \alpha' X_t = \alpha' X_0 + \alpha' C^*(L) \varepsilon_t$$

To get the common trend representation, one needs to work on the algebra of X_t representation:

$$X_t = X_0 + C(1) \left[\tilde{\mu}_t + \sum_{s=1}^t \varepsilon_s \right] + C^*(L) \varepsilon_t$$

where $\tilde{\mu}_t = C(1)^{-1} \mu_t$

$$X_t = X_0 + C(1) H \left[H^{-1} \tilde{\mu}_t + H^{-1} \sum_{s=1}^t \varepsilon_s \right] + a_t$$

where $a_t = C^*(L) \varepsilon_t$

$$X_t = X_0 + S_k^{-1} C(1) H \left[S_k H^{-1} \tilde{\mu}_t + S_k H^{-1} \sum_{s=1}^t \varepsilon_s \right] + a_t$$

And finally,

$$X_t = X_0 + A\tau_t + a_t$$

where $S_k^{-1}C(1)H = A$, $\tau_t = \pi + \tau_{t-1} + u_t$, $\tau_t = S_k H^{-1} \mu_t + S_k H^{-1} \sum_{s=1}^t \varepsilon_s$ and $a_t = C^*(L)\varepsilon_t$.

Jong (2002) remarks this same point, as we explore in the third subsection. The $I(1)$ term of a Beveridge decomposition corresponds to $I(1)$ term on a Stock Watson decomposition, since the latter one is a multivariate extension for cointegrated systems of the first, an one univariate framework.

Following the authors, the benefit from their decomposition, compared to Stock and Watson, is that it is easier to estimate and test hypotheses on the common long-memory components.

On the estimation procedures ⁷, Gonzalo and Granger (1995) estimate $\gamma\alpha'$ from (2.14) by regressing ΔX_t and X_{t-1} on $(\Delta X_{t-1}, \dots, \Delta X_{t-q+1})$. By doing this, they are able to get the residuals of the regressions and then build another equation to estimate $\gamma\alpha'$, as below.

$$R_{0t} = \gamma\alpha' R_{1t} + \epsilon_t \tag{2.24}$$

The parameter $\gamma\alpha'$ is estimated by reduced rank regression. After solving the eigenvalues problem and finding the maximum likelihood estimators, they use the maximum likelihood function to choose the estimator of γ which gives the maximum value for the likelihood function. They also present alternatives ways to estimate γ , considering that α was not esti-

⁷They base their proofs on Johansen (1988).

mated by simultaneous reduced rank least square or maximum likelihood. They point out that the main outcome from this methodology is the possibility of testing different linear combinations of X_t in order to check if it is in fact a common factor.

In the empirical part of the paper, they present three examples. The first and second one (based on the work of Cochrane (1991)) is a GNP (gross national product) and consumption case, followed by a dividends and stock prices case, where they show how to obtain the common factors directly from the error correction model. They find that if consumption is fixed, only the transitory component can affect GNP. In the second example, they find that a shock in dividends is characterized as permanent. However, a shock in prices (keeping dividends constant) has only transitory effects. In the third example, they apply their methodology, by decomposing the common factors in permanent and transitory, using interest rate data from Canada and United States. They find that there is only one common factor among six different interest rates, which turns to be the U.S common factor.

Many authors follow this methodology and also change the estimation in some aspects, such as Booth, So, and Tseh (1999), Chu, Hsieh, and Tse (1999) and Harris et al (2000). Some suggest to estimate A_1 in (2.15) (or B_1 in (2.16)) as a measure of price discovery, leading to the component share approach. We discuss their work in the empirical literature review section. Some other authors work in order to compare this methodology with the one of Hasbrouck (1995). In the next two subsections, we first see Hasbrouck (1995) proposal and then we compare the two methodologies, finalizing with the empirical findings.

2.2.2 Information Share (IS)

Hasbrouck (1995) introduces a new methodology to measure price discovery among different markets. He points out two approaches currently used in the literature to model these markets: common implicit efficient price (as already mentioned from the work of Garbade and Silber (1979) and Garbade and Silber (1983)) and “lead and lag” returns regressions. The argument used by Hasbrouck (1995) in order to use the same approach as Garbade and Silber (1979) and Garbade and Silber (1983) and not the lead and lag return regressions, comes from the fact that the latter one is commonly misspecified, from an econometric viewpoint. Hence, Hasbrouck (1995) defines price discovery as how it is measured by this approach:

“Price discovery in this framework refers to innovations in the efficient price. A market’s contribution to price discovery is its information share, defined as the proportion of the efficient price innovation variance that can be attributed to that market.” Hasbrouck (1995), page 1177

Following the definition above, all prices of a same security (or derived security) have the same component in its price structure within all different markets the security is traded, which is the common unobserved efficient price.

The beginning of this methodology explanation relies on the concept of cointegration. All equations and explanations below are from Hasbrouck (1995), otherwise, it is specified.

Considering two different markets trading the same security, its prices

are defined as:

$$p_{1,t} = p_{1,t-1} + w_t \quad (2.25)$$

$$p_{2,t} = p_{1,t-2} + \varepsilon_t \quad (2.26)$$

w_t and ε_t are zero mean *i.i.d* variables. Both prices are non stationary and a linear combination of the two forms a stationary variable, which means they are cointegrated.

$$p_{1,t} - p_{2,t} = p_{1,t} - (p_{1,t-2} + \varepsilon_t) = p_{1,t} - (p_{1,t-1} - w_{t-1}) - \varepsilon_t = w_t + w_{t-1} - \varepsilon_t \quad (2.27)$$

Rewriting (2.25) and (2.26) in terms of price changes, we have a Vector Moving Average representation (VMA) only in terms of the errors components, as stated in (2.28).

$$\Delta p_{1,t} = w_t \quad (2.28)$$

$$\Delta p_{2,t} = p_{2,t} - p_{2,t-1} = p_{1,t-2} + \varepsilon_t - (p_{1,t-3} + \varepsilon_{t-1}) \quad (2.29)$$

$$\Delta p_{2,t} = \Delta p_{1,t-2} + \varepsilon_t - \varepsilon_{t-1} = w_{t-2} + \varepsilon_t - \varepsilon_{t-1} \quad (2.30)$$

One may show that both equations have a common component. First, using (2.25) and rearranging terms, we have for $p_{1,t}$ and $p_{2,t}$:

$$p_{1,t} = p_{1,0} + \left(\sum_{s=1}^t w_s \right) \quad (2.31)$$

$$\Delta p_{2,t} = p_{2,t} - p_{2,t-1}$$

$$p_{2,t} = \Delta p_{2,t} + p_{2,t-1} \quad (2.32)$$

Now, substituting (2.29) into (2.32):

$$p_{2,t} = w_{t-2} + \varepsilon_t - \varepsilon_{t-1} + p_{2,t-1} \quad (2.33)$$

Building the same equation for the lagged variable, delivers:

$$p_{2,t-1} = w_{t-3} + \varepsilon_{t-1} - \varepsilon_{t-2} + p_{2,t-2} \quad (2.34)$$

Substituting (2.34) into (2.33):

$$p_{2,t} = w_{t-2} + \varepsilon_t - \varepsilon_{t-1} + w_{t-3} + \varepsilon_{t-1} - \varepsilon_{t-2} + p_{2,t-2}$$

By doing this recursively, the ε_{t-1} terms cancel each other, and we end up with the equation below, where (as in (2.31)), the price is written as the sum of an *i.i.d* variable $\left(\sum_{s=1}^t w_s\right)$, which is common to both prices ($p_{1,t}$ and $p_{2,t}$).

$$p_{2,t} = p_{2,0} + \left(\sum_{s=1}^t w_s\right) - w_t - w_{t-1} + \varepsilon_t \quad (2.35)$$

This common term to both prices is later viewed as the implicit efficient price, considering some additional conditions. As mentioned before, an alternative way to model price changes is using lead and lags. By substituting

recursively (2.29):

$$\begin{aligned}
\Delta p_{2,t} &= \Delta p_{1,t-2} + \varepsilon_t - \varepsilon_{t-1} \\
\Delta p_{2,t-1} &= \Delta p_{1,t-3} + \varepsilon_{t-1} - \varepsilon_{t-2} \\
\varepsilon_{t-1} &= \Delta p_{2,t-1} - \Delta p_{1,t-3} + \varepsilon_{t-2} \\
\Delta p_{2,t} &= \Delta p_{1,t-2} + \varepsilon_t - \Delta p_{2,t-1} + \Delta p_{1,t-3} - \varepsilon_{t-2} \\
&\dots
\end{aligned} \tag{2.36}$$

One realizes it is necessary to have infinite lags in order to specify it correctly, which does not converge. Another way of writing this model is called error correction model (from (2.25) and (2.26)):

$$\begin{aligned}
\Delta p_{1,t} &= w_t \Delta p_{2,t} = p_{2,t} - p_{2,t-1} \\
\Delta p_{2,t} &= (p_{1,t-2} + \varepsilon_t) - p_{2,t-1} \\
p_{1,t-2} &= p_{1,t-1} - \Delta p_{1,t-1} \\
\Delta p_{2,t} &= (p_{1,t-1} - p_{2,t-1}) - \Delta p_{1,t-1} + \varepsilon_t
\end{aligned} \tag{2.37}$$

The economic interpretation for this last expression comes from the fact that traders in market number 2 react looking at the price difference between the two markets ($p_{1,t-1} - p_{2,t-1}$) in period $t - 1$ minus any shock that happens in market number 1 in period $t - 1$ ($\Delta p_{1,t-1}$) plus any shock from market 1 in period t .⁸

⁸This last expression may be written in many other ways, such as:

$$\Delta p_{2,t} = (p_{1,t-2} - p_{2,t-2}) - \Delta p_{2,t-1} + \varepsilon_t$$

or even

$$\Delta p_{2,t} = (p_{1,t-2} - p_{2,t-2}) + \varepsilon_t$$

Considering n price variables integrated of order one (price changes are covariance stationary), containing a random walk component and all linked to a single security, their vector moving average (VMA) expression is (as defined in (2.28) and (2.29)):

$$\Delta p_t = \Psi(L)e_t \quad (2.38)$$

where e_t is a zero mean vector with covariance matrix equal to Ω . The vector e_t is also assumed to be serially uncorrelated. $\Psi(L)$ is a lag polynomial. Also consider the following equation:

$$\beta'_{(n-1) \times n} = [\iota_{n-1} : -I_{n-1}].^9 \quad (2.39)$$

By multiplying this matrix by p_t series, the resulting expression ($\beta'p_t$) is stationary (since it is a first difference of prices). Also, considering the polynomial $\Psi(1)$ as the sum of the moving average coefficients, as $\beta'p_t$ is stationary, this leads to $\beta'\Psi(1) = 0$. All rows of $\Psi(1)$ need to be identical for this to be true given the structure of β' . As $\Psi(1)e_t$ is the sum of the differences between a disturbance term in period t and a disturbance term in period $t-1$, it represents intuitively the disturbances long run impact, which are common to all prices, since the rows of Ψ are equal to each other. We can write the equation to describe prices, considering the $\Psi(1) = \psi p_0$ a constant vector ($n \times 1$) and $\Psi^*(L)$ a matrix polynomial in the lag operator. This results comes by applying Beveridge-Nelson decomposition to (2.38).¹⁰

¹⁰Lutkepohl (2007) explains the Beveridge-Nelson decomposition. Considering an $I(1)$ process y_t , with a stationary first difference being written as an infinite MA representation, as below

$$\Delta y_t = w_t = \theta(L)u_t$$

$$p_t = p_0 + \psi \left(\sum_{s=1}^t e_s \right) \iota + \Psi^*(L)e_t \quad (2.40)$$

Where $\psi \left(\sum_{s=1}^t e_s \right)$ is a scalar, that, when multiplied by a unit column (ι) it turns to be the same for all prices. $\Psi^*(L)e_t$ is a zero mean covariance stationary process.

Prices can be written as below:

$$A(L)p_t = k + e_t \quad (2.41)$$

$$A(L) = I - A_1L - A_2L^2 - \dots - A_kL^k \quad (2.42)$$

Finally, as already defined in (2.13), an error correction model may be written in the form below:

$$\Delta p_t = \alpha(\beta' p_{t-1} - E\beta' p_t) + \Gamma_1 \Delta p_{t-1} + \Gamma_2 \Delta p_{t-2} + \dots + \Gamma_{k-1} \Delta p_{t-k+1} + e_t \quad (2.43)$$

where the moving average coefficients satisfy the conditions:

$$\sum_{j=0}^{\infty} j |\theta_j| < \infty$$

$$\theta(1) = \sum_{j=0}^{\infty} \theta_j \neq 0$$

$$u_t \sim (0, \sigma_u^2)$$

where u_t is a white noise. Beveridge-Nelson decomposition shows that such a process can be written as the sum of a random walk ($\theta(1)(u_1 + \dots + u_t)$), a stationary process ($\sum_{j=0}^{\infty} \theta_j^* u_{t-j}$) and initial values ($y_0 - w_0^*$), as below:

$$y_t = y_0 + w_1 + \dots + w_t = y_0 + \theta(1)(u_1 + \dots + u_t) + \sum_{j=0}^{\infty} \theta_j^* u_{t-j} - w_0^*$$

The correspondence between (2.43) and (2.41) is stated below:

$$\alpha\beta' = -A(1)$$

$$\Gamma_j = -\sum_{i=j+1}^K A_i$$

for $j=1, \dots, K-1$.

A few notes on how to estimate (2.43) comes first on the determination of the VMA order in (2.38) which is done by iterating forward (2.41) and taking the first difference. In (2.43), the first estimative is on the $E\beta'p_t$ term (sample average) and after this, the equation can be estimated using linear least squares.

From (2.40), the term $\psi \left(\sum_{s=1}^t e_s \right)$ is common to all prices, which may be assumed to be due to new information. Given that e_t has a covariance matrix equal to Ω , the variance of $\psi \left(\sum_{s=1}^t e_s \right)$ is equal to $\psi\Omega\psi'$. Now, if we consider that p_t is the price in n different markets, and $e_{j,t}$ is the innovation term in the j^{th} market, $\psi\Omega\psi'$ has n terms (considering Ω diagonal) and each of these terms corresponds to the innovation term of each of the n markets. Hasbrouck (1995) calls the market j 's information share the proportion between the innovation form market j ($\psi_j^2\Omega_{jj}$) and the total innovation ($\psi\Omega\psi'$).

$$S_j = \frac{\psi_j^2\Omega_{jj}}{\psi\Omega\psi'} \quad (2.44)$$

If Ω is not diagonal, price innovations are correlated across markets and then the information share cannot be computed from the ratio in (2.44). The solution for this problem comes from two different approaches. The first and simplest one would be to use very short intervals between observa-

tions (high frequency data), since we expect that much of this correlation among markets come from contemporaneous effects. A change in a market happens and a few seconds later, another market reflects this same change, in a sequential way. However, this methodology is not always effective, as it may only reduce, but not eliminate the correlation among markets. The second approach involves the triangularization of the covariance matrix, in order to determine upper and lower bounds. Following that, one may impose that innovations in n markets are given by

$$e_t = Fz_t$$

Where z_t are random variables ($Ez_t = 0, Var(z_t) = I$) and F is a Cholesky factorization¹¹ of Ω ($\Omega = FF'$). In this case, the information

¹¹Given a MA (∞) representation as:

$$y_t = \mu + \varepsilon_t + \Psi_1\varepsilon_{t-1} + \Psi_2\varepsilon_{t-2} + \dots$$

Considering a variance-covariance matrix of ε_t being equal to Ω . As Ω is symmetric positive definite matrix, it can be written in the form:

$$\Omega = ADA'$$

here A is a unique lower triangular matrix with 1s on the principal diagonal and D is a unique diagonal matrix with positive entries on the principal diagonal. By using matrix A , one is able to build a vector u_t as below:

$$u_t = A^{-1}\varepsilon_t$$

$$E(u_t u_t') = D$$

The Cholesky Decomposition of a matrix Ω is given by:

$$\Omega = AD^{1/2}D^{1/2}A' = PP'$$

where

$$P = AD^{1/2}$$

D is a diagonal matrix, where the (j, j) element is the variance of u_{jt} . $D^{1/2}$ is the diagonal matrix whose (j, j) element is the standard deviation of u_{jt} . P is a lower triangular matrix, just like A . However, instead of having 1s on the principal diagonal, as A , P presents the standard deviation of u_t along its principal diagonal. The explanation

share would be calculated by:

$$S_j = \frac{([\psi F]_j)^2}{\psi \Omega \psi'} \quad (2.45)$$

In such a case, he ends up considering upper and lower bounds for the information share. By implementing Cholesky factorization two times (in the case of two markets), when market 1 is the first variable in the factorization, then it is the upper bound, when it is the second variable, it is the lower bound.

Hasbrouck (1995) also applies this methodology to compute their empirical results. He uses data set from NYSE and regional exchanges in the USA and he found that price discovery happens mainly on NYSE, with the median information share of NYSE being equal to 92.7 percent.

2.2.3 Comparison: Permanent Temporary (PT) and Information Share (IS)

Following these two main methodologies to measure price discovery among different markets, considerable work has been done on which would be the most appropriate approach. This subsection highlights some comparisons and comments on Gonzalo and Granger (1995) and Hasbrouck (1995) work.

Baillie, Booth, Tse, and Zobotina (2002) make a theoretical comparison between the two approaches and an empirical application. According to their work, both approaches start from the same model, a vector error

above comes from Hamilton (1994) pages 318-323.

correction model, as below,

$$\Delta Y_t = \alpha\beta'Y_{t-1} + \sum_{j=1}^k A_j \Delta Y_{t-j} + e_t \quad (2.46)$$

with the following covariance matrix (Ω) of innovations (e_t):

$$\Omega = \begin{pmatrix} \sigma_1^2 & \rho\sigma_1\sigma_2 \\ \rho\sigma_1\sigma_2 & \sigma_2^2 \end{pmatrix}$$

The $\alpha\beta'Y_{t-1}$ in (2.46) is the long run component, which represents the long run relation between the two prices, whereas $\sum_{j=1}^k A_j \Delta Y_{t-j}$ describes short run movements from market imperfections.

Up to here, both methodologies are the same. From here on, they start to go apart according to Baillie, Booth, Tse, and Zobotina (2002).

Hasbrouck (1995) rewrites equation (2.46) using a vector moving average representation, as in (2.38). Then, he transforms it into its integrated form, where one is able to see the common efficient price component $\left(\psi \left(\sum_{s=1}^t e_s \right) \right)$, which is a sum of shocks on prices coming from new information (by assumption).

Gonzalo and Granger (1995), differently, define the common factor component as a linear combination of Y_t , as in equation (2.16). They prove that the error correction term does not Granger cause the common factor in the long run, as well as introduce a methodology to test if a single factor component is the main responsible for the common factor.

Baillie, Booth, Tse, and Zobotina (2002) points out that while Hasbrouck (1995) decomposes the common factor innovations variance ($\psi\Omega\psi'$),

Gonzalo and Granger (1995) decompose the common factor as a linear combination of the two prices, as below.

$$f_t = b_1 x_{1t} + b_2 x_{2t} \quad (2.47)$$

They also demonstrate that what really matters is the relative values of ψ_j and γ_j and prove the following

$$\frac{\psi_1}{\psi_2} = \frac{b_1}{b_2}. \quad (2.48)$$

Substituting (2.48) into (2.44), one ends up with:

$$S_j = \frac{b_j^2 \Omega_{jj}}{b_1^2 \Omega_{11} + b_2^2 \Omega_{22}} \quad (2.49)$$

$$\frac{S_1}{S_2} = \frac{b_1^2 \Omega_{11}}{b_2^2 \Omega_{22}} \quad (2.50)$$

Baillie, Booth, Tse, and Zabotina (2002) say that if Ω_{11} and Ω_{22} are similar, results using the Information share methodology and the Permanent Temporary methodology bring similar results.

As said in the previous subsection, these two equations are only valid considering that the error terms are not correlated. When this does not hold, Hasbrouck (1995) proposal is to use Cholesky factorization to overcome the problem, and in this situation, he works with lower and upper bounds for the information share. Following (2.45), matrix F can be written

as

$$F = \begin{pmatrix} f_{11} & 0 \\ f_{12} & f_{22} \end{pmatrix} = \begin{pmatrix} \sigma 1 & 0 \\ \rho \sigma_2 & \sigma_2(1 - \rho^2)^{1/2} \end{pmatrix}$$

Considering (2.45) and (2.48):

$$\frac{S_1}{S_2} = \frac{(b_1 f_{11} + b_2 f_{12})^2}{(b_2 f_{22})^2} \quad (2.51)$$

Considering the information share in each of the markets, given $S_1 + S_2 = 1$ (by construction),

$$\begin{aligned} S_1 &= \frac{(b_1 f_{11} + b_2 f_{12})^2}{(b_1 f_{11} + b_2 f_{12})^2 + (b_2 f_{22})^2} \\ S_2 &= \frac{(b_2 f_{22})^2}{(b_1 f_{11} + b_2 f_{12})^2 + (b_2 f_{22})^2} \end{aligned} \quad (2.52)$$

The venue first factorized has a higher S , as shown in (2.52), since one more term ($b_2 f_{12}$) is considered, which is the upper bound. When one implements this factorization in the opposite way (market 2 first), then one finds market's upper bound, and market's 1 lower bound.

Hasbrouck (1995) finds lower and upper bounds very close to each other (by using small intervals between observations, contemporaneous correlation becomes insignificant), however, as Baillie, Booth, Tse, and Zobotina (2002) remark, some other studies find significant differences on the bounds, making results interpretation difficult. They also point out that equations (2.51) and (2.52) can be easily estimated using directly VECM, instead of VMA, which is used by Hasbrouck (1995) and many others.

Baillie, Booth, Tse, and Zobotina (2002) perform some empirical appli-

cation in order to compare the two methodologies. They compare prices in ECNs ¹², wholesalers, wire houses, institutional brokers and others. The results from the information share methodology delivers that ECNs contribute 58.6% of the innovations on the efficient price. The other four groups are responsible for the remaining 41.4%. By using the PT methodology, they find that ECNs comprise 69.3% of the efficient price, and the other 4 market makers contribute for 30.7%. Therefore, both approaches conclude that the ECNs dominate the other four groups in price discovery. They point out that the reason why IS brings a smaller price discovery share than PT is that while the IS model incorporates the correlation between the series, the PT does not.

As a conclusion, according to Baillie, Booth, Tse, and Zobotina (2002), the IS and the PT define price discovery in a different way. In one hand, IS considers price discovery in terms of the variance of the innovations to the common factor (as from (2.44)), making the market that has the greatest share of the volatility to have the greater share of the price discovery (by assuming that price volatility reflects the flow of information). On the other hand, PT considers each market's contribution to the common factor, decomposing this last one itself. They do not indicate one methodology as being better than other, however they agree that Hasbrouck's approach has more general economic appeal and interpretation.

Hasbrouck (2002) briefly explains the two methodologies, where he points out an important drawback of Gonzalo and Granger (1995)'s methodology. The traded price is defined as a common factor plus a stationary process, however, the common factor does not have the random walk prop-

¹²ECN is a computer market that disseminates buy and sell limit orders, acting as quasi stock exchange and as a broker.

erty implied. From this, Hasbrouck (2002) states that once the common factor is not a random walk, it is also not a martingale. From his words:

“If it is not a random walk, then it cannot be a martingale, nor therefore can it be an unbiased conditional expectation of the security’s eventual value. Nor will its variance generally equal the long run variance of the security price. One cannot go so far as to say that such a component could never be of interest to a trader (or econometrician), but a justification could only be based on the particulars of well defined structural model. Outside such a model, it is difficult to conjecture why a permanent non-martingale price component warrants general interest.” Hasbrouck (2002) page 332

He also comes up with some application using three structural models. He gets better results on the market price discovery share by using the information share methodology than by using the permanent transitory approach in two of the three models. In the other one, he gets the correct share using PT, however, the variance and first order autocorrelation of efficient price change are overestimated. In this same example, IS methodology gives a wide range on the estimated share (although it has the correct value in the range), but computes correct estimates of variance and first order autocorrelation of efficient price change. He concludes that the benefit of the PT approach is that it may achieve a precise estimation, when the IS approach presents a significant difference between the lower and upper bound (although containing the correct value); however, the PT estimates are more volatile and autocorrelated than the efficient price. Following that, the IS methodology supports more meaningful inference.

Harris, McInish, and Wood (2002) also compare the two methodologies on their work. They use practical estimation (same simulated data as Hasbrouck (2002)) to show that Gonzalo and Granger (1995)'s methodology brings the true information structure in many financial market microstructure models and also can be used in statistical testing, while the Hasbrouck's methodology cannot.

They conclude that Gonzalo and Granger's methodology provides useful and attractive method for recovering the information structure of the asset pricing models used by Hasbrouck (2002).

Jong (2002) writes a brief review on each methodology and shows that the random walk part of Gonzalo and Granger permanent component is equal to the Stock-Watson common factor (which is also stated on the prepositions from the first work). He substitutes the Stock-Watson decomposition into Gonzalo and Granger linear combination of the common factor and X_t . Below, the first equation refers to Stock-Watson decomposition, the second only shows Gonzalo and Granger assumption (the permanent component being a linear function of X_t , as in (2.16)) and the third equation shows them combined.

$$X_t = \alpha_{\perp} \theta' \sum_{s=0}^t \varepsilon_s + C^*(L) \varepsilon_t \quad (2.53)$$

$$f_t = B_1' X_t = (\gamma_{\perp}' \alpha_{\perp})^{-1} \gamma_{\perp}' X_t \quad (2.54)$$

$$f_t = (\gamma_{\perp}' \alpha_{\perp})^{-1} \gamma_{\perp}' (\alpha_{\perp} \theta' \sum_{s=0}^t \varepsilon_s + C^*(L) \varepsilon_t) \quad (2.55)$$

$$f_t = \theta' \sum_{s=0}^t \varepsilon_s + s_t \quad (2.56)$$

Jong (2002) concludes that Information Share and Permanent and Tran-

sitory methodology are closely related. Given the Stock Watson decomposition, in one hand, Information Share looks at a normalized $\theta_i^2 \sigma_i^2$ term, measuring the contribution of ε_i compared to the total variance of the innovation in the permanent component. On the other hand, Gonzalo and Granger look at a normalized θ_i , measuring the impact of ε_i on the innovation of the permanent component. Hence, both of them are normalized versions of the term θ , which is expressed in Hasbrouck efficient price (once he uses Stock-Watson decomposition) equation as well as on Gonzalo and Granger permanent component. Jong (2002) concludes that both methodologies have their merits, although, Hasbrouck's definition is a more proper way to measure the information amount generated in each market.

Lehmann (2002) writes a summary of Baillie, Booth, Tse, and Zobotina (2002), Jong (2002), Hasbrouck (2002) and Harris, McInish, and Wood (2002) papers, all based on the comparison between Hasbrouck (1995) and Gonzalo and Granger (1995) methodology. He concludes, by enumerating differences and similarities from the work of the above, as below, in his words:

- the Gonzalo-Granger portfolio weights do not resolve the inherent identification problem associated with the measurement of the contribution of different markets to price discovery but rather are proportional, to the long run impact multiplier

$$\lim_{\tau \rightarrow \infty} \partial p_{t+\tau} / \partial \varepsilon'_t$$

- hence, the Stock-Watson common trend should be used to estimate efficient price;

- these portfolio weights generically identify markets in which all price discovery takes place in large samples only when the Hasbrouck information shares do so as well; and

-in a number of examples involving uncorrelated trade innovations across markets in which all price discovery takes place in one market but neither method attributes all price discovery to it with the Gonzalo-Granger weights declining sharply as the trading costs in the second market and, thus, its trade innovation variance shrink. Lehmann (2002), page 276

Yan and Zivot (2010) propose the use of a structural cointegration model¹³, in order to show what the results for the two price discovery measures (IS and component share (CS) (the applied version of Gonzalo and Granger methodology)). They state that a clear comprehension of price discovery is only possible by using a structural model (Structural Moving Average), as below:

$$\Delta \mathbf{p}_t = \mathbf{D}(L)\eta_t = \mathbf{D}_0\eta_t + \mathbf{D}_1\eta_{t-1} + \mathbf{D}_2\eta_{t-2} \dots \quad (2.57)$$

Where $D(L) = \sum_{k=0}^{\infty} D_k L^k$, $D_0 \neq I_2$, D_0 and $D(L)$ are invertible, $\eta_t = (\eta_t^P, \eta_t^T)'$, being the permanent and transitory shock, respectively; which are uncorrelated and have a diagonal covariance matrix. Equation (2.57) can be rewritten as:

$$\begin{pmatrix} \Delta p_{1t} \\ \Delta p_{2t} \end{pmatrix} = \begin{pmatrix} d_1^P(L) & d_1^T(L) \\ d_2^P(L) & d_2^T(L) \end{pmatrix} \begin{pmatrix} \eta_t^P \\ \eta_t^T \end{pmatrix}$$

¹³The model they propose is motivated by structural VAR models, as they point out, which are very much used in empirical macroeconomic exercises.

Where $d_1^P(L)$, $d_1^T(L)$, $d_2^P(L)$, $d_2^T(L)$ are lag polynomials tracing the responses of the permanent and transitory impacts on prices. Two defining characteristics must be pointed out on permanent and transitory shocks, as below:¹⁴

$$\lim_{k \rightarrow \infty} \frac{\partial E_t [p_{t+k}]}{\partial \eta_t^P} = \lim_{k \rightarrow \infty} \sum_{l=0}^k \frac{\partial E_t [p_{t+l}]}{\partial \eta_t^P} = \lim_{k \rightarrow \infty} \sum_{l=0}^k D_l^P = D^P(1) = 1 \quad (2.58)$$

$$\lim_{k \rightarrow \infty} \frac{\partial E_t [p_{t+k}]}{\partial \eta_t^T} = \lim_{k \rightarrow \infty} \sum_{l=0}^k \frac{\partial E_t [p_{t+l}]}{\partial \eta_t^T} = \lim_{k \rightarrow \infty} \sum_{l=0}^k D_l^T = D^T(1) = 0 \quad (2.59)$$

By following these conditions, matrix $D(1)$ can be written as

$$D(1) = \begin{bmatrix} 1 & 0 \\ 1 & 0 \end{bmatrix}$$

Hence, (2.57) may be derived in a different way, once one applies Beveridge-Nelson decomposition:

$$\mathbf{p}_t = \mathbf{p}_0 + D(1) \sum_{j=1}^t \eta_j + \mathbf{s}_t = \mathbf{p}_0 + \mathbf{1m}_t + \mathbf{s}_t \quad (2.60)$$

where $\mathbf{s}_t = D^*(L)\eta_t \sim I(0)$, $D_k^* = -\sum_{j=k+1}^{\infty} D_j$, $k=0, \dots, \infty$ and $\mathbf{m}_t = \mathbf{m}_{t-1} + \eta_t^P$.

The $D(1)$ matrix represents the long run impact (the integrated part of the equation), while the \mathbf{s}_t term is integrated of order zero and represents any deviation from the efficient price. The last part of equation (2.60) comes from Stock and Watson (1988) suggestion.

By rewriting (2.40) from the Information Share methodology and considering the integrated of order 1 term as common to all prices (since there exist a the vector ι) equal to the efficient price m_t , one have the following

¹⁴As remarked by the authors, a similar condition was made by Gonzalo and Granger (1995) as it is reproduced in equation (2.17).

expression:

$$\mathbf{p}_t = \mathbf{p}_0 + \Psi(1) \left(\sum_{s=1}^t e_s \right) + \Psi^*(L) \mathbf{e}_t = \mathbf{p}_0 + \mathbf{1} \mathbf{m}_t + \mathbf{s}_t \quad (2.61)$$

With $\mathbf{s}_t = \Psi^*(L) \mathbf{e}_t$ and $\mathbf{m}_t = \mathbf{m}_{t-1} + \eta_t^p$, where η_t^p is the permanent chock, once it affects the efficient price. By doing this, Yan and Zivot (2010) remark that the only difference between the two different methodologies, which are stated by equations ((2.60) and (2.61)) is that the transitory shocks (\mathbf{s}_t) in the first equation are driven by the structural innovation (η_t) instead of \mathbf{e}_t . As the authors point out, this implies that pricing errors in the short run may contain the permanent term (η_t^p), which means that some liquidity effect of information-related trading and the lagged price adjustment to new information may be included in this term. According to the authors, the parameters in (2.57) can be estimated using the reduced form of the VEC model, as in (2.43), and applying a modification of the permanent and transitory decomposition of Gonzalo and Ng (2001). Yan and Zivot (2007) demonstrate the estimation procedure one may apply to get the parameters in (2.57).

Yan and Zivot (2010) analyze the relation between the structural model and the IS and CS measures. From (2.38)¹⁵ and (2.57), we have that $\mathbf{e}_t = \mathbf{D}_0 \eta_t$, which leads to the following equations when there are two markets:

$$e_{1,t} = d_{0,1}^P \eta_t^P + d_{0,1}^T \eta_t^T \quad (2.62)$$

$$e_{2,t} = d_{0,2}^P \eta_t^P + d_{0,2}^T \eta_t^T \quad (2.63)$$

¹⁵Considering $\Delta p_t = \Psi(L) \mathbf{e}_t = \mathbf{e}_t + \Psi_1 \mathbf{e}_{t-1} + \Psi_2 \mathbf{e}_{t-2} + \dots$

As \mathbf{D}_0 is invertible, one may write η_t^P and η_t^T as a function of the errors in the reduced form.

$$\begin{aligned}\eta_t^P &= \frac{d_{0,2}^T}{\Delta} e_{1,t} - \frac{d_{0,1}^T}{\Delta} e_{2,t} \\ \eta_t^T &= -\frac{d_{0,2}^P}{\Delta} e_{1,t} - \frac{d_{0,1}^P}{\Delta} e_{2,t}\end{aligned}\tag{2.64}$$

Where $\Delta = |\det \mathbf{D}_0|$. From Hasbrouck (1995), Yan and Zivot (2010) define¹⁶:

$$\eta_t^P = \psi' e_t = \psi_1 e_{1,t} - \psi_2 e_{2,t}\tag{2.65}$$

Where ψ_1 and ψ_2 are the parameters used in the information share measures, as in (2.44). Hence, by (2.64) and (2.65):

$$\begin{aligned}\psi_1 &= \frac{d_{0,2}^T}{\Delta} \\ \psi_2 &= -\frac{d_{0,1}^T}{\Delta}\end{aligned}\tag{2.66}$$

These equations show that the parameters used in order to measure information share are proportional to the market's contemporaneous responses to the structural frictional innovation, as Yan and Zivot (2010) reveals.

The authors also derive the component share formula using the structural model. In the next chapter, we present the extension of Gonzalo and Granger Permanent and Transitory methodology to the empirical application Component Share (by Booth, So, and Tseh (1999) and Chu, Hsieh,

¹⁶On (2.40), it is implied that the permanent component, which is common to all prices, is equal to $\Psi(1) \left(\sum_{s=1}^t e_s \right) \iota$.

and Tse (1999)), which is defined as ¹⁷:

$$CS_i = \frac{\alpha_{\perp,i}}{\alpha_{\perp,1} + \alpha_{\perp,2}}, i = 1, 2$$

Yan and Zivot (2010) show that one may write the component share measure using the parameters defined to estimate the information share, as below¹⁸:

$$CS_i = \frac{\psi_i}{\psi_1 + \psi_2}, i = 1, 2 \quad (2.67)$$

This representation illustrates the fact that the IS measure is a variance weighted version of the component share, when market innovations are uncorrelated.

As in (2.66) we find the parameters ψ_i with relation to the matrix \mathbf{D}_0 , one can substitute them into the CS measure in (2.67). The results for the case of two markets are stated below:

$$CS_1 = \frac{d_{0,2}^T}{d_{0,2}^T - d_{0,1}^T} \quad (2.68)$$

$$CS_2 = \frac{-d_{0,1}^T}{d_{0,2}^T - d_{0,1}^T} \quad (2.69)$$

These equations show that the component share measure only includes the transitory parameter.

The authors also present some structural models examples and conclude that both measures (Information Share and Component Share) alone cannot quantify price discovery, since the first one includes permanent and

¹⁷From (2.14), (2.15) and (2.16)

¹⁸Baillie, Booth, Tse, and Zabotina (2002) and Jong (2002) have also notice that, however not applying to the CS measure.

temporary shocks, while the second one has only the transitory component. In order to sort this issue out, the authors propose a different measure, in order to have a correct specification of the relative impact of permanent shocks:

$$\left| \frac{IS_1 CS_2}{IS_2 CS_1} \right| = \left| \frac{d_{0,1}^P}{d_{0,2}^P} \right|$$

Jong and Schotman (2010) propose a new measure of price discovery, extending the univariate case of Hasbrouck (1993) to a multivariate case. They propose a structural form of a first order vector moving average process (VMA), as below:

$$p_t = \iota p_t^* + u_t \tag{2.70}$$

$$p_t^* = p_{t-1}^* + r_t \tag{2.71}$$

$$u_t = \alpha r_t + e_t \tag{2.72}$$

Where $Var(r_t) = \sigma^2$ and $Var(e_t) = \Omega$.

They find the moments conditions for Δp_t , where all parameters are identified with exception of α . To find that, they use the reduced form of the model, a VMA process, which is given by:

$$\Delta p_t = \epsilon_t - C\epsilon_{t-1}$$

where $Var(\epsilon_t) = \Sigma$ and $C = I - \iota\theta'$.

They use Beveridge and Nelson representation on the reduced form of the model, and then are able to write the relation between the reduced

form and the structural form, given by:

$$Cov(\Delta p_t, r_t) = \Sigma \theta = \sigma^2(\iota + \alpha)$$

This expression gives a particular choice for the parameter α . Other ways to identify α include to set the α of one market equal to zero (generalization of Watson restriction, where all α 's are equal to zero), which means that the idiosyncratic term is uncorrelated with the efficient price, hence, that this market is the central market. Another alternative is to assume Ω diagonal, which is very different then imposing Σ , as the authors point out, since the latter one can be easily violated in empirical applications. To propose a new measure for price discovery, they define price innovations from the structural model as:

$$v_t = \iota r_t + u_t = (\iota + \alpha)r_t + e_t = \beta r_t + e_t$$

The covariance matrix is given by:

$$E[v_t v_t'] = \Upsilon = \sigma^2 \beta \beta' + \Omega$$

Where the innovation in the efficient price is equal to shocks in the individual prices plus a term which is unrelated to innovations in observed prices, as below:

$$r_t = \gamma' v_t + \eta_t$$

In the structural model (Unobserved Components), the variance of η_t is positive, while in the reduced form, the term is always equal to zero, by

construction. The regression coefficients are:

$$\gamma = \Upsilon^{-1}\beta\sigma^2$$

The variance of r_t is given by:

$$Var(r_t) = \sigma^2 = \gamma'\Upsilon\gamma + \sigma_\eta^2$$

They calculate the total fraction of variance in the fundamental price innovation, which is explained by the observed price innovations with the equation below:

$$R^2 = 1 - \sigma_\eta^2/\sigma^2 = \gamma'\Upsilon\gamma/\sigma^2 = \gamma'\beta = \sum_{j=1}^N \gamma_j\beta_j$$

The authors propose the term $\gamma_j\beta_j$ as a measure of price discovery (information share).

Finally, assuming Ω diagonal with positive diagonal elements ω_j^2 , and given the definition above of information share, one has:

$$IS_j = \frac{\beta_j/\omega_j^2}{1/\sigma^2 + \sum \beta_i^2/\omega_i^2}$$

This measure is exactly the same as the one of Hasbrouck (1995) when Σ is diagonal.

Their empirical results (application to a set of Nasdaq dealer quotes) show a comparison between a VECM, a VMA and the structural model (unobserved components model), where the results from the first two models present a wide range between minimum and maximum values for the

information share. On the structural model, they present three results: the one with Watson restriction, the one with Ω being diagonal, and the one where Ω is approximately diagonal. They find that diagonal appears to be a good modeling assumption and that the unobserved components models can deliver parsimonious results and more informative estimates of information share than the reduced form models.

Another alternative methodology comes from Grammig and Peter (2012) who propose to solve the main drawback of Hasbrouck's approach to measure price discovery: the fact that in most empirical cases, the methodology delivers upper and lower bounds, instead of a unique measure. Their contribution aims to identify a unique measure, given the distributional properties of financial data, such as fat tails (large negative or positive price changes happen more often than predicted by normal distribution) and tail dependence (correlation of price changes in the tails is different than in the rest of the distribution). They connect these two facts with the insights that when the data exhibit heteroscedasticity, it can be described by multi-regimes processes associated with different innovation variances, leading to the possibility of identifying structural innovations in a multiple time series framework. They use these data characteristics to disentangle the contemporaneous correlations of the price innovations.

2.3 Empirical Literature Review

Apart from the literature on price discovery methodology and econometrics issues, there is a large literature on empirical findings. In the past years, the increase of cross listings in many trading venues and the increment in

accessability of data sets led to a rich research environment in price discovery. This section presents the relevant empirical application on this topic. We start by briefly mentioning a different group of the price discovery literature that does not employ high frequency data. If one uses high frequency data, price discovery can be inferred using a range of methodologies, such the ones we presented in the previous section. Our focus is in this group of research. The interest of the former group is not which market incorporates news first, but understand other aspects of cross listing. Gagnon and Karolyi (2010), for instance, analyze price deviations from parity and their comovements with market indexes and currencies, whereas Karolyi (2004) studies how cross listed firms impact market development. Doidge, Karolyi, and Stulz (2001) measure cross listing impact on companies market value and Baruch, Karolyi, and Lemmon (2007) study traded volume allocation of cross listed firms among home and U.S market. Foerster and Karolyi (1999) analyze how returns change from previous period of cross listing to post period and show that listing decreases the local market risk exposure, but it does not significantly change the exposure to global market risk. We start by giving details on this low frequency literature and then we move towards our main interest, high frequency data.

2.3.1 Low Frequency Data

Gagnon and Karolyi (2010) work with a ADR (American Depositary Receipt) and the corresponding stock traded in the domestic market. Their sample include companies from all over the world, but European and Canadian companies are majority, with 41% and 24% of their sample size, respectively. They study price deviations from parity and their comovements

with market indexes and currencies. By the no-arbitrage condition, they assume that in an integrated equity market the price differential between ADR and its domestic share should be equal to zero, once one adjust for exchange rate. They propose a log differential on USA index and home market index in order to measure the comovement of stock prices over time.

$$R_{A-H,t} = \alpha + \sum_{i=-1}^{+1} \beta_i^{US} R_{M,t+i}^{US} + \sum_{i=-1}^{+1} \beta_i^H R_{M,t+i}^H + \sum_{i=-1}^{+1} \beta_i^{FX} R_{FX,t+i} + \varepsilon_{A-H,t}$$

Where $R_{A-H,t}$ is the difference of the log price differential, $P_t^A - P_t^H$ between the ADR price (P_t^A) in US dollars and home market price in US dollars (P_t^H). $R_{M,t+i}^{US}$ is the US return, while $R_{M,t+i}^H$ is the home market return, and the log currency change of the benchmark currency for the home market relative to US market is $R_{FX,t+i}$. They work with daily data and find significant price differences among stocks traded simultaneously at different venues, however, they rarely persist for more than one day. Also, returns differentials exhibit comovements relative to market index returns on their respective trading location.

Karolyi (2004) analyses whether companies that start to cross list in the US stock market facilitate or hinder the development of home stock market. He evaluates a wide range of measures of stock market development, such as ratio of market capitalization to GDP, number of publicly listed firms, overall cross border equity flows and trading activity, analysing these measures at the firm level, split into into ADR firms and non ADR firms. Also, he distinguishes between the different types of ADR listings available in the US market (the ones listed in major stock exchanges and the ones

listed in smaller markets). The data set includes six emerging markets in Latin America and in Asia and varies where it begins in 1976 in the earlier cases and in 1989 in the latter ones and ends in 2000. In order to measure stock market development, he develops four indicators: market capitalization ratio (value of listed shares divided by GDP), number of publicly traded companies, turnover ratio (value of total shares traded divided by market capitalization and capital flow ratio (total value of monthly portfolio equity flows over GDP). On the regression analysis of these indicators, the investigation is on the influence of the ADR activity on the stock market development proxies, by using cross-sectional time-series regression on a multi-country and multivariate analyses. The equation on development indicators (y_{it}) is below:

$$y_{it} = \alpha_i + x'_{it}\beta + z'_{it}\gamma + \delta_i y_{it-1} + \varepsilon_{it}$$

Where x'_{it} is ADR activity variables¹⁹, and z'_{it} is a number of other factors that may influence development. There is one equation for each development indicator and the regression allows for country fixed effects by including a country parameter (α_i). The results show that the growth of ADR programmes in emerging markets facilitate the expansion of cross border equity flows in those countries (although uneven between countries). Companies that are not cross listed in the US market suffer from negative spillovers, since their capitalization and turnover ratios decline, once the

¹⁹It includes: the fraction of total number of stocks included in the IFCG Global index for each market which have ADR listed in the US, the fraction of US dollar value trading in the IFCG global index for each market that is comprised of trading in ADRs listed in the US and the fraction of the US dollar market capitalization of the IFCG Global index for each market that is represented by the market capitalization of the ADRs listed in US.

number of companies in ADR programmes increase. He also finds that the negative spillovers are statistically and economically large even for smaller ADR listings when considering different types of ADR.

Doidge, Karolyi, and Stulz (2001) analyze the market value of cross listed companies compared to non cross listed companies from the same country. They find that foreign companies listed in the US market have a Tobin's q ²⁰ 16.5% higher than companies from the same country not listed (result from the end of 1997). They run some regressions to examine the effect of listing on Tobin's q and to check if their theory of listing positively impacting firms value is correct. Tobin's q is given by the equation below:

$$q_i = \alpha + \beta' X_i + \delta L_i + \varepsilon_i$$

Where X_i is a set of exogenous country variables, L_i (decision on listing) is a dummy variable which is equal to 1, if the company is cross listed, and zero, otherwise. Since L_i and ε_i are correlated, they assume L_i is given by:

$$L_i^* = \gamma' Z_i + \eta_i \tag{2.73}$$

$$L_i = 1 \text{ if } L_i^* > 0 \tag{2.74}$$

$$L_i = 0 \text{ if } L_i^* < 0 \tag{2.75}$$

Where L_i^* is an unobserved latent variable, Z_i is a set of variables that influence the listing decision (L_i). They use Heckman's two step estimator for the two equations above. They find a cross-listing premium, whose persistence is still significant even when they control for country and firm

²⁰Tobin's q ratio is given by equity market value plus liabilities from book value divided by total assets (equity book value +liabilities). Hence, if $q > 1$, market value is higher than book value, and if $q < 1$, book value is higher than market value.

specific characteristics. They present a literature review on cost and benefits of cross listing and propose a theory in order to explain why cross listed companies are valued more. They conclude that firms choosing to be cross listed are firms with a better alignment between shareholders and controlling shareholder's interests. This is because controlling shareholders of companies cross listed in the US cannot extract as many private benefits from control, compared to companies not cross listed. There are a number of SEC reporting and compliance requirements, which makes private benefits opportunities to appear less often. Also, cross listed firms are valued more because a smaller part of their cash flow is expropriated by controlling shareholders and because cross listed firms are more able to take advantage of growth opportunities.

Baruch, Karolyi, and Lemmon (2007) are more interested in volume traded than in value. They develop a theoretical model of multi-market trading to explain the variation in the US share of global trading volume for firms cross listed in the U.S. Under certain conditions their model predicts that the distribution of trading volume between stock exchanges is related to the correlation between cross listed asset's return and other assets traded in this stock exchange returns. Their database is composed by 275 firms cross-listed in NYSE/Nasdaq from 25 emerging and developed countries, with monthly data on prices and volume from January 1995 to December 2001. They try to prove that when the return of the cross listed asset is more sensitive to information in the U.S. market relative to information in the domestic markets, the U.S market has a higher share of the asset's overall trading volume. They perform a variance decomposition of returns to understand the contribution of each market's index return on the asset's

return. Firstly, the two equations below are estimated:

$$R_{i,t} = \alpha_i + \beta_{i,H}R_{Home,t} + \epsilon_{i,t} \quad (2.76)$$

$$R_{i,t} = \alpha_i + \beta_{i,H}R_{Home,t} + \beta_{i,US}R_{US,t} + \epsilon_{i,t} \quad (2.77)$$

Where $R_{i,t}$ is the return for stock i , $R_{Home,t}$ is the market index's return in the home country and $R_{US,t}$ is the return on the U.S market index. Considering the first equation as the restricted regression, and the second one as unrestricted, they compute a F-statistic for each stock (assuming n monthly observations), in order to capture the incremental contribution of U.S market movements in explaining price variation compared to the firm's home market. The F-statistic, or U.S information measure, is below:

$$\frac{(R_{UR}^2 - R_R^2)/2}{(1 - R_{UR}^2)/(n - k - 1)}$$

They also control for factors that may have an effect on the U.S. share of trading volume at the country level. They run multivariate regressions with this measure as the independent variable (or its natural log), plus the natural log of the firm's market value of equity, an indicator variable for cross listed firms in Nasdaq/Amex and an indicator for emerging markets or country fixed effects, and the U.S. share of traded volume (or its natural log) for each firm as the dependent variable. They find the coefficient on the U.S. information measure statistically and economically significant, as it shows it has an effect that supports the predictions of their theoretical model, that is, volume migrates to the stock exchange in which the cross listed asset returns presents a higher correlation with market index's return.

Mei, Scheinkman, and Xiong (2005) build three empirical hypothesis that are tested with Chinese stock market data, relating speculative behaviour in dual-class shares (same company, two class of shares) and price difference between these two shares. The first empirical hypothesis is that there is a positive relationship between the speculative component in asset prices and the turnover of shares. The second says that when investors are risk averse, have different beliefs and cannot short sell, the speculative component in share prices and the share turnover rate decrease with asset float. The last one, states that when investors trade purely for liquidity reasons, the turnover rate of shares increases with asset float. They work with two daily price series, one from class A shares (restricted to domestic residents) and another from class B share (for foreigners), from 1993 to 2001. The same company issues both type of shares. The firm's A share price is given by

$$P_{it}^A = \frac{E_i}{R_{it}^A - g_i} + S_{it}^A$$

Where $\frac{E_i}{R_{it}^A - g_i}$ is the fundamental component of prices (current expected value of discounted future dividends adjusted for risk premium, with E_i being the expectation of current earnings, g_i is its growth rate and R_{it}^A is the discount rate) and S_{it}^A is the speculative component, which depends on the volatility of the difference in beliefs among Chinese investors about the firm's fundamentals value and other factors. The firm's B share is:

$$P_{it}^B = \frac{E_i}{R_{it}^B - g_i} + S_{it}^B$$

Firm's A and B share premium can be expressed as

$$\rho_{it} = \frac{P_{it}^A - P_{it}^B}{P_{it}^B} = \frac{R_{it}^B - g_i}{R_{it}^A - g_i} + \frac{S_{it}^A}{S_{it}^B} - 1$$

They run the cross-sectional regression below to analyse the variation in A-B share premia, with monthly average turnover rates.

$$\rho_{it} = c_{0t} + c_{1t}\tau_{it}^A + c_{2t}\tau_{it}^B + \epsilon_{it}$$

Where $\tau_{it}^A = \log(1 + turnover_{it}^A)$ and $\tau_{it}^B = \log(1 + turnover_{it}^B)$. They find that the price difference between A and B shares is positively related to the turnover rate of A shares. They also control for some variables, such as liquidity, risk premium and discount rates, when they still find the same results. They find that this price difference increases with firm's idiosyncratic return volatility and decreases with the float of A shares.

Hence, by using the foreign share prices to control for variations in firm's fundamentals (as both shares have the same rights), they find results indicating the existence of a speculative component in the prices of domestic shares and that speculative trading is an important determinant of stock prices in bubbles.

Foerster and Karolyi (1999) study how returns change from the before cross listing period to after period. Their sample is composed by 153 listings of 11 countries with daily data. They present a summary of previous studies in the subject, concluding that global cross-listing may lead to a reduction in expected return if the capital markets where the security is traded is partially or completely segmented. They first estimate each firm's return relative to market level return (α and β) using data from day -250 to -101

(pre listing), and they find an abnormal return estimation, using returns from day -100 to +250, as below:

$$\epsilon_{it} = R_{it} - [\alpha_i + \beta_{iL} R_{mt}^L]$$

Where R_{it} is firm i 's return on day t . They find strong evidence of an increase in price preannouncement. Average abnormal returns during days -100 and -2 are 0.11 percent, while around the announcement period (days -1 and 0), the average abnormal return increases to 0.21 percent. After announcement, average returns are not significantly different than zero. They also present this analysis for weekly returns, showing that on the weeks before listing, returns increase by 0.38 percent per week, which is statistically significant. After listing, returns present an average decline of 0.27 percent. The authors explain this price behaviour around listing/announcements days by stating the hypothesis already existent in the literature that return should be positive and greater for firms in which the home market is more segmented from the US market and smaller for firms in which the home market more integrated to the US market. They also use a modified version of IAPM (International Asset Pricing Model), adjusting for domestic and global risk. They run the equation below:

$$R_{it} = \alpha_i^{PRE} + \beta_{iL}^{PRE} R_{mt}^L + \beta_{iW}^{PRE} R_{mt}^W + \alpha_i^{LIST} D_t^{LIST} + \alpha_i^{POST} D_{it}^{POST} + \\ \beta_{iL}^{POST} R_{mt}^L D_{it}^{POST} + \beta_{iW}^{POST} R_{mt}^W D_{it}^{POST} + \epsilon_{it}$$

Where α_i 's are constant and seen as abnormal excess returns. β_{iL} 's are the coefficients on the local market excess return (R_{mt}^L). β_{iW} 's are the coefficients for global market index excess return (R_{mt}^W), D_{it} 's are dummy

variables, which are equal to 1 (if observations are from listing week or from post listing period), and 0 otherwise. They find that the average local β in the pre listing period is 1.03 and the global is much smaller, equal to 0.22. However, the post listing β 's is equal to 0.74 (local) and 0.12 (global), but the change in the global coefficient is not significant. These findings show that listing decreases the local market risk exposure, but it does not significantly change the exposure to global market risk.

2.3.2 High Frequency Data

We now turn our attention to empirical applications using high frequency data. The empirical research focuses is on finding the role of each venue in contributing to price formation. Some papers discuss price discovery across exchanges. Others, develop an across markets analysis, such as derivatives markets. A third group carries out the study of different platforms, such as electronic and floor trade. Some authors study price discovery across hours, comparing trading during open market hours and after hours.

Harris, McNish, and Wood (2002) use Gonzalo and Granger (1995) methodology in order to measure the common factor weight of three markets (New York, Chicago and Pacific) for the 30 DJIA²¹ stocks for three years (1988, 1992 and 1995). First, they test for cointegration vectors and find that trading prices in the three exchanges follow an error correction process with the cointegrating vector being α , as defined below:

$$P_t = A_1 f_t + A_2 z_t = A_1 \gamma_{\perp} + A_2 \alpha P_{t-1}$$

²¹Dow Jones Industrial Average

They estimate the Gonzalo and Granger (1995) factor weights ($\gamma_{\perp i}$) in order to understand what is the contribution of each exchange on the single common factor²². The New York average weight was 72% in 1988, declining to 52% in 1992 and 63% in 1995, while Pacific presented weights of 15%, 21% and 15%, and Chicago, 13%, 28% and 22%, in 1988, 1992 and 1995, respectively. Then, they use Gonzalo and Granger (1995)'s Q_{GG} statistic to test if 100% of price discovery happens at NYSE as the null hypothesis, by saying that the weight is equal to one for NYSE and zero to the other two exchanges. They reject the null hypothesis for six stocks in 1988, for nineteen stocks in 1992 and for eleven in 1995, showing that the regional stock exchanges have important role on the common factor formation. The results they find in 1988 for many stocks are consistent with the results that Hasbrouck (1995) finds that NYSE is “information dominant”. They present results for five alternative data collection procedures (REPLACE ALL, MINSPAN and XFIRST, X=NYSE, Pacific and Chicago).

Chu, Hsieh, and Tse (1999) study the price dynamics in three S&P 500 index markets: the S&P 500 spot index, future contracts on S&P 500 index and a new index product, Standard and Poor's Depositary Receipts (SPDRs)²³. They first analyse the cointegration behaviour of these three price series, by estimating a vector error correction model with matched synchronous intraday data from 1993 (they use Harris, McInish, Shoesmith, and Wood (1995) technique in order to identify the tuples and make the data synchronous). They find that the three indices are cointegrated with

²²There is only one common factor, they work with three markets ($n=3$) and two cointegrating vectors ($r=2$).

²³SPDRs are exchange-traded securities that represent ownership in the SPDR Trust. The SPDR trust is a long term unit trust the intends to track the price performance and dividend yield of the S&P 500 index.

two cointegrating vectors. They also decompose the common factor in the cointegration system, by using Gonzalo and Granger (1995) methodology. The representation is given by:

$$Y_t = \theta f_t + \tilde{Y}_t \quad (2.78)$$

$$f_t = \alpha'_\perp Y_t \quad (2.79)$$

$$\theta = \beta_\perp (\alpha'_\perp \beta_\perp)^{-1} \quad (2.80)$$

They find all three markets contribute to the price discovery process by using the χ^2 test of Gonzalo and Granger. Also, the results show that the futures market is the dominant component of the common stochastic trend, followed by spot and then SPDRs (they find $\alpha'_\perp = 1.0, 0.144$ and 0.222 , for futures, spot and SPDRs, respectively).

Booth, So, and Tseh (1999) try to understand whether index options contribute to the price discovery process involving index securities, by using German DAX²⁴ stock index, index futures and index options. To solve the problem of nonsynchronicity on the data, they employ two techniques from Harris, McInish, Shoesmith, and Wood (1995): REPLACE ALL and MINISPAN. They use a VECM to describe the relation between the three series returns and Gonzalo and Granger (1995) methodology to express the common factor as a linear combination of the price vector, normalizing the parameters in order to have the parameters in percent terms and hence, express the parameters as a common factor weight. They find that the contribution on the common factor is given 50% by the DAX index, 48% by the future DAX (FDAX) and only 2% by options on DAX (ODAX). The

²⁴DAX is a German index (formerly Deutscher Aktien-Index) on the Frankfurt Stock Exchange and consists of the thirty major German companies.

χ^2 also delivers the same result, that price discovery is shared equally by DAX index and FDAX. This goes in line with their hypothesis that price discovery happens in the market in which there are the smallest overall transaction costs, once the DAX futures appear to have the lowest trading costs, and the DAX options, the highest.

Harris, McInish, Shoesmith, and Wood (1995) find that equilibrium price for IBM stock occurs not only from information from NYSE, but also from Midwest exchange and Pacific exchange. They use data from these three stock exchanges for IBM to estimate an error correction model to understand the contribution of each stock exchange on price discovery. Their hypothesis is that if prices from Midwest and Pacific exchanges respond to price deviations from NYSE and prices in NYSE do not respond to price deviations in Pacific and Midwest exchanges, this means that price discovery is focused in NYSE. The authors point out that in this application, the error correction model involves only cross market information flows (adjustments to price disparity across the three markets), and not innovations in IBM prices due to new information revelations. They show that each of the three exchanges make an independent and significant impact on IBM prices on the other two exchanges, although NYSE prices react less to price variations in the other exchanges. They present results from REPLACE ALL, MINSPAN and REPLACE OLDEST procedures.

Tse, Bandyopadhyay, and Shen (2006) study the dynamics of price in the Dow Jones Industrial Average (DJIA) and in its three derivative products: DIAMOND exchange-traded fund ETF (which is traded just like a stock), floor traded regular futures and the electronically traded mini futures (sized at one-half of the regular futures contracts to make accessible

to small investors). They work with intraday data from May until July of 2004. By applying the Hasbrouck (1995) information share methodology, the authors find the contribution of different ECN's on the price discovery of DIAMOND ETF trades. ArcaEx is shown to account for 59% of the price discovery, while Island accounted for 28.3%, Amex for 5.5%, NYSE for 3.7% and Nasdaq for 3.5%. Secondly, the authors apply the same methodology, but now trying to understand the price discovery percentage component of DIAMOND traded in the non electronic exchange Amex (trade and quotes, which is an average between bid and ask), futures on DJIA and E-mini futures in DJIA on the determination of DJIA index. They find that the E-mini futures almost dominates the other markets, since it contributes for 96.2% of price discovery. They do the same exercise using DIAMOND traded at ArcaEx (electronic trading), instead of DIAMOND traded at Amex and find that the contribution of E-mini drops to 69%, resulting in ArcaEx quotes playing an important role, as it contributes with 28.6%. The authors insert a five second delay to the reported transaction time for E-mini future prices and find that ArcaEx quotes (80.3%) now dominates E-mini futures (12%). They implement a robustness check by using S&P 500 and its derivatives. These results indicate that multi-market trading contributes to a better pricing efficiency and that informed traders are attracted by the anonymity and immediate service provided by electronic trading platforms of the E-mini futures and exchange-traded funds.

Hasbrouck (2003) works with three U.S. equity indexes (S&P 500, S&P 400 and Nasdaq-100) traded in ETFs (exchange-traded index mutual funds), in floor traded and electronically-traded index futures contracts to estimate

the price discovery participation of each of these markets. By using Hasbrouck (1995) methodology and high frequency data to estimate the information share, he finds that in the indexes where there are E-mini contracts (S&P 500 and Nasdaq-100), they provide the largest information share. In the S&P 400 index, the information share is higher on the ETF.

Theissen (2005) analyses the price discovery process in two different types of trading system: the floor-based and the electronic exchange. The study expects to find the electronic trading system impounding new information faster into prices, since electronic systems are less costly to operate (hence, present a lower bid ask spread) and more liquid (there more traders, given the remote access). Also, as most existing electronic trading systems are anonymous, it is likely that this type of system attracts more the informed trader, where he can exploit this advantage. The study is undertaken with data from German stock market over the months of June and July, 2007, containing time-stamped transaction prices, trading volumes and best bid and ask quotes from both trading systems of the thirty stocks in the index DAX. Theissen uses two methodologies (Information Share and Gonzalo and Granger factor weights) in order to estimate the contribution of each market on the price discovery process. The system's contribution is almost the same when using transactions prices, while when using quotes midpoints, the electronic trading systems present a more important role, which is consistent with the higher quotation activity on this system. The average between the lower and upper bounds from Hasbrouck (1995) methodology and the factor weights presented similar results on his analysis for most of the stocks and the results confirm the finding that both systems contribute to the price discovery process. When considering the

mean between all the stocks on the two methodologies, a *t-test* is run with a null hypothesis of equal means for the two methodologies, that is, both measuring the same price discovery contribution the markets analysed. The null hypothesis cannot be rejected at 5% significance for three models (1-minute and 5-minute interval using transaction prices and 1-minute interval using quote midpoints) and is strongly rejected for one model (5-minute interval using quote midpoints)²⁵. The study also shows a cross-sectional analyses in order to confirm that the price discovery contribution of each market is positively related to their market shares and concludes that floor trading is not necessarily inferior to electronic system considering the aspects of price discovery contribution and that the more liquid stocks have appeared to show more the advantages of electronic trading.

Martens (1998) studies how the type of market is influential to the price discovery role of each market. He works with Bunds futures contracts (long term German Government Bonds) which are very liquid in both markets where are traded at, in London (LIFFE) in a floor trading system and in Germany (DTB) in an electronic system. Given that the presence of a limit market book (one of the main difference between the two systems is that the electronic includes it) brings an advantage in very quiet periods (market depth) and a disadvantage in fast moving markets (slackness of changing prices), it is expected to be found that the contribution to the price discovery process is larger in fast moving markets than in quiet periods for outcry markets (non electronic system). He breaks the sample into periods of high and low volatility and estimates information share (Hasbrouck (1995) methodology is used) for both markets in these two cat-

²⁵According to the author, the information to estimate this model is the least precise.

egories of volatility. The empirical results confirm his hypothesis. In high volatility periods the floor trading system has the largest share in the price discovery process, while in low volatility periods, the electronic system is the one to present the highest contribution. He concludes that the two systems should be viewed not only as competitors, but as complement to each other, suggesting an hybrid trading system.

Grammig, Melvin, and Schlag (2005) use data from NYSE and Frankfurt (XETRA) of three German firms to analyse where the price discovery process occurs and also how these prices respond to exchange rates shocks. They use an adaptation of Hasbrouck (1995), once their model contains two common trends, one as in Hasbrouck being the efficient stock price, and the second one as being the efficient exchange rate. The period analysed is from August, 1999 to October, 1999, bid - ask midpoint quote and the interval data is 10 seconds, chosen since lower frequency would bring more evidence of contemporaneous correlation and higher frequency would not bring significant gains on this issue, but it would bring some sources of microstructure “noise”. In summary, they show the importance of a three-variable model and find the exchange rate to be exogenous with respect to the two stock prices for the three stocks, that the exchange rate shocks affect more the prices at NYSE than in Frankfurt and that most of the price fundamental component is determined in Frankfurt. Although they find for one company an information share for NYSE equal to 20% (for the other two firms, the percentage is lower than that), the results strongly suggest that for a company cross listed internationally, the home market is the one to play the main role on the price discovery process.

Barclay and Herdershott (2003) analyse how trade during the day or after the trading hours affect incorporation of news into prices. They expect to find trading differences between the normal period of trade (from 9:30AM to 4:00PM) and after and before this period (they concentrate the analysis from 8:00AM to 9:30AM and from 4:00PM to 6:30PM). Considering that the information asymmetry declines over the trading period and that the cost of holding a suboptimal portfolio overnight may be higher, one expects to find a larger number of liquidity-motivated trades in the post-close and a higher fraction of informed trades in the pre open. They find that after hours, low volume can generate significant price discovery, but prices during this period are noisier, which may imply that price discovery is less efficient.

Jong (1998) study price discovery in foreign exchange market. They use data from October, 1992 to September, 1993 to try to prove their hypothesis that German banks are price leaders in the Deutschmark/dollar market. They use a methodology related to Hasbrouck (1995), although, instead of a VECM model, they work with a structural time series model. They find that some banks do present a higher information share than others, among this group some large German banks seem to have a more important role on the price discovery process, however, German banks are not exclusive in this group.

Chapter 3

Dealing with a High Frequency Data Set

3.1 Institutional Background

The use of data from the Brazilian stock exchange is particularly interesting. The large time intersection between Brazil and US provides just a very small daily period of time where information is coming only from one market, because of opening and closure hours. The two markets have a time overlap period of six hours and a half during the majority of the year, from mid February to mid November, with the Brazilian exchange being open only for thirty minutes while NYSE is closed. The smallest intersection between the two markets occurs from mid November to mid February, where they are still both open for five hours and a half. This is of great value to analyse the price discovery dynamics, since markets are less likely to lose their importance on price discovery because of trading hours. Hence, it becomes easier to isolate the different aspects driving the

price dynamics in all markets. This is in stark contrast with price discovery analyses that employ European stocks and their ADR counterparts: Due to time difference, the intersection is of only from 2 to 3 trading hours. Figure 3.1 shows the time intersection between the Brazilian and US markets during the year in Brazilian time. These six hours and thirty minutes intersection on the majority of the year deliver a significant advantage for the price discovery analysis compared to other studies in the literature.

The BM&FBovespa is the only stock exchange in Brazil and the leading exchange in Latin America in terms of number of contracts traded. In 2002, Bovespa bought equity membership of the Rio de Janeiro stock exchange (BVRJ) and Bovespa merged with the Brazilian Mercantile & Futures Exchange (BM&F), forming the BM&FBovespa in 2008. It is a fully electronic exchange (end of open outcry transactions at Bovespa was in 2005 and derivatives transactions was in 2009) and operates under supervision of the CVM (Brazilian Securities Commission). BM&FBovespa markets include equity, commodities and futures, foreign exchange, securities and ETF's (exchange traded funds). BM&FBovespa presented a market capitalization of USD 1.2 trillion in 2012, not too far from many European stock exchanges, such as Deutsche Borse (USD 1.5 trillion), BME Spanish (USD 1.0 trillion) and SIX Swiss (USD 1.2 trillion) being one of largest stock exchanges in the world (top 13 in market capitalization). The London Stock Exchange presents a market capitalization of USD 3.64 trillion

Brazil achieved the investment grade rating from Standard & Poor's in April 2008. Fitch and Moody's increased Brazilian rating in May 2008 and September 2009, respectively. IBOVESPA is the most important index at BM&FBovespa, the index and the exchange as a whole have reflected

the improvement in these ratings in the past five years with a trading volume increase of 7.3% (compound annual growth rate). The IBOVESPA index is composed by the most traded stocks, being the main indicator of the Brazilian stock market's average performance. The 62 companies composing the IBOVESPA present a combined market value of USD 1.242 trillion, as in the end of December 2010.

The data set from Bovespa ranges from December 2007 to November 2009 and covers the entire of trades from all stocks, with price, quantity and time as Table 3.1 shows. There are 834 securities traded and 442 listed companies.

Tables 3.2 and 3.3 show basic features of the 20 most traded companies for 2008 and 2009, respectively. This is one of the main explanations for the firms we chose to work with, as we explain in more details in the next section. Table 3.4 presents how transactions on Bovespa behave around the opening time of NYSE. We show the opening hours of Bovespa (B) and NYSE (N) during the first and second quarters of 2008. We chose these two quarters just to illustrate that the number of trades at Bovespa increase around the opening hours of NYSE.

Foreign companies usually are traded and listed in the US market through American Depositary Receipts (ADR). An ADR is a physical certificate evidencing ownership of a US dollar denominated form of equity in a foreign company. It represents the shares of the company held on deposit by a custodian bank in the company's home country and carries the corporate and economic rights of the foreign shares, subject to the terms specified on the ADR certificate.

3.2 Data Description

We use a tick by tick data set of Brazilian blue-chip companies traded at three different venues, Bovespa, NYSE and ARCA. The sample period is beneficial since it is large enough to englobe a variety of movements in the stock markets, including the 2008/2009 financial crises. With this data set, we measure price discovery considering stable and highly instable periods. The use of a high frequency data set provides timely incorporation of new information in each different market. A daily data set of these markets would not provide the information needed to measure price discovery, since at a day to day level, all markets would have incorporated all new information.

In the next chapter we focus on the two most liquid stocks in the BM&FBovespa, namely, Petrobras and Vale. They are both constituents of the IBOVESPA, the main benchmark indicator of the Brazilian capital markets. Petrobras is a publicly-traded integrated oil and gas multinational, whose main stockholder is the Brazilian government with over 55% of the common shares. It is the fifth largest energy company in the world, with presence in 28 countries. It performs as an energy company in the following sectors: exploration and production of oil and gas in offshore fields, refining, oil and natural gas trade and transportation, petrochemicals, and derivatives, electric energy, biofuel among others. It is a leader in the Brazilian oil industry, and aims to be among the top five integrated energy companies in the world by 2020.

Vale was founded as a public company by the Brazilian Government, being privatized in 1997. It is the second largest metals and mining company in the world and the biggest private sector company in Latin America.

Vale has a market capitalization of around USD 160 billion. Vale is the world's largest producer of iron ore and iron ore pellets, where the majority of their revenues come from. It is also the world's second largest producer of nickel, beyond their production of manganese, ferroalloys, thermal and coking coal, copper, cobalt, platinum group metals and fertilizer nutrients. They also have operations on energy, logistics and steel, but not as being its core business. As a result from Vale's recent diversification strategy, the participation of non-ferrous metals (notably, nickel, copper, and kaolin) on total revenues has recently increased in a substantial manner.

Petrobras and Vale issue both common and preferred shares at the BM&FBovespa. In addition, they are also present at the NYSE through the ADR program at the highest level a foreign company may sponsor (i.e., level 3, allowing for listing and public offering). Petrobras and Vale are the most active ADRs in the NYSE, both by trading value and volume. The ADRs respond for about 30% of the Petrobras outstanding shares (26% for commons and 34% for preferreds), whereas these figures for Vale are about 25% for common shares and 40% for preferred shares. Our data set includes the prices of both common and preferred shares of Petrobras and Vale in Brazil as well as their ADR prices in the US from January 2008 to November 2009. This gives way to a system of 5 market prices for Vale: exchange rate, common and preferred share prices in Brazil and in US. For Petrobras, we are also able to distinguish trades at the NYSE from transaction at the NYSE Arca,¹ leading to a system of 7 market prices.

In the last chapter, we extend our analysis to more firms, including a

¹ The NYSE Arca exchange in Chicago is the second largest electronic communication network in terms of shares traded. It results from a reverse merge on February 27, 2006 between the NYSE Group and Archipelago Holdings.

wide variety of industries in our study. Brazilian companies are very liquid in the US market, sometimes having more trading activity there than in Brazil. We choose to use very liquid companies in the three markets, in order not to lose information once aggregating a very illiquid firm with one presenting a much larger number of trades. We work with the following firms: Ambev (beverage), BR Telecom (telecommunication), Bradesco (finance), Gerdau (steel), Vale (mining) and Petrobras (oil). Apart from BR Telecom, they all belong to IBOVESPA. Preferred shares of Vale and Petrobras are the two most heavily traded shares at the Brazilian market, with Gerdau and Bradesco coming in the top 15. The number of trades at Bovespa for preferred Petrobras is around 9 million for the 2-year data set. For Vale is 6.4 million, Gerdau 3.2 million, Bradesco 3 million, Ambev 0.7 million and 0.6 for BR Telecom. At Arca and Nyse, these stocks also present a significant number of trades, being in some cases larger than the volume traded at the Brazilian exchange (see table 3.6).

Figure 3.2 shows the price evolution for the shares used in this study. There are periods of large price instability, characterized specially by the effects of the 2008/2009 financial crises. A significant drop in prices is observed for the majority of stocks, jointly with an increase in volatility.

3.3 Data cleaning and aggregation

Given that the goal is to check how timely markets react to news incorporating them into prices, it is paramount to work with intraday data. Sampling data at a lower frequency could well blur all sorts of lead-lag patterns between different assets and/or trading platforms. Suppose for

instance that we employ daily data and trading platform B is less liquid than trading platform A. In the presence of new information, prices in A would react on average more quickly than prices in B, but they would both converge to the same fundamental value in the long run (i.e., as soon as enough transactions hit both trading platforms). As a matter of fact, it is very likely that this convergence takes place before market close at least for actively traded assets. The use of daily data would completely miss price B lagging a few seconds or minutes behind price A due to the proximity of the closing prices.

On the other hand, employing tick data raises a number of data handling issues. To control for reporting errors and delays as well as, to some extent, for microstructure effects (e.g., bid-ask bounce), we first purge the data from observations that seem implausible not only given the usual market conditions, but also given the market activity at the time. In particular, as in Brownlees and Gallo (2006), we exclude any price that does not satisfy $|p_i - \bar{p}_i(k, \delta)| < 3s_i(k) + \gamma$, where $\bar{p}_i(k, \delta)$ and $s_i(k)$ are respectively the δ -trimmed sample mean and the sample standard deviation of a neighborhood of k observations around i , and γ is a granularity parameter to avoid zero variances from a sequence of k equal prices. We restrict attention to neighborhoods within the same trading day. For instance, the first k prices of the day compose the neighborhood of the first observation, whereas the last k prices of the day form the neighborhood of the last observation. However, in general, neighborhoods are given by the first preceding $k/2$ prices and the following $k/2$ prices.

The above discriminant aims to validate observations on the basis of how much they deviate from what we expect given a neighborhood of valid

observations. This means one should choose the filter parameters very carefully. The trimming parameter δ should obviously increase with the frequency of outliers, whereas k should increase with trading intensity. It turns out that the filter is much more sensitive to changes in γ than in the other parameters and so we set the granularity to the minimum price variation of 0.01. We fix δ at 10% and specify k according to the number of trades, ranging from 20 to 60 observations. As a robustness check, we construct alternative data sets by varying the values of (k, γ, δ) . Tables 3.5 and 3.6 report the initial number of observations and the number of outliers we discard for each price series as well as the resulting sample sizes after the filtering.

The next step is to deal with the nonsynchronicity of tick data. Table 3.5 documents that common shares have much more ticks, and so more liquidity, than preferred shares in the US, especially for the electronic Arca platform. In contrast, preferred shares are much more actively traded than common shares at the BM&FBovespa. The reason for this combination of common shares high concentration and preferred shares high circulation in Brazil is mainly historical. First, the Brazilian government revoked in 1997 the article of the Brazilian Corporate Act that granted tag-along rights to common shareholders in order to promote the privatization program. As a consequence, common shares became much less appealing, with liquidity further migrating towards preferred shares. Second, Brazilian firms could issue two preferred shares for each common share until 2001, enabling shareholders to increase their capital leverage without diluting power. Although the ratio is now one to one for new issues, the overall ratio still causes imbalances between political and economic power, increasing the

possibility of wealth expropriation.

Although it is possible to examine price discovery in tick time, Frijns and Schotman (2009), we take the traditional route by aggregating data into regular intervals of time. This allows using the standard VECM/VMA machinery that permeates Hasbrouck’s (1995) information share framework. As for the sampling frequency, the literature documents a trade-off between market microstructure noise and contemporaneous correlation between markets. As the data frequency increases, microstructure effects become more apparent, whereas the contemporaneous correlation presumably declines. As the spectral-based IS measure is robust to contemporaneous correlation, we give more weight to alleviating market microstructure effects as what concerns the choice of the sampling frequency. In particular, we sample prices at intervals of 30 and 60 seconds by capturing the most recent trade on each market.

Tables 3.7 and 3.8 show the number of observations before and after the aggregation procedure. As expected, liquidity is a chief concern for common shares in Brazil (namely, Petr3 and Vale3) due to their low circulation. The low trade intensity leads to many missing observations due to the absence of trades even at the 30-second frequency. This could lead to spurious serial correlation and hence we employ the Newey-West covariance matrix estimator in the analysis. As a robustness check, we estimate the covariance matrix using different lag structures (including no lags) as well as consider 60-second intervals in order to reduce the fraction of zero returns. The results are qualitatively very similar and hence we omit them to conserve on space. Needless to say, they are available upon request.

It is interesting to notice that there are less intervals with zero returns

in the US market than at the home market. The latter seems sufficiently liquid only for preferred shares, whereas the proportion of zero returns are much more reasonable for the NYSE. We show in the next section that these liquidity concerns indeed matter, playing a major role in the price discovery analysis.

For the last chapter, we aggregate these series based on a defined time interval ranging from 30s to 300s, as some stocks are more intensively traded than others at certain periods and/or overall the time span. We allow this range because some stocks do not present enough trades to aggregate at a higher frequency, for instance 30s. This is the case for BR Telecom. Aggregating at a higher frequency would result in a large amount of missing observations (since some 30s-intervals would not have any trades), which may lead to serial correlation. The benefits for the price discovery measure would not be that high, since no trades are happening. Even though we are careful on choosing the intervals for each firm, we estimate the covariance matrix using the Newey-West estimator in order to control for serial autocorrelation. To aggregate the series we use the methodology proposed by Harris, McNish, Shoesmith, and Wood (1995). For each interval, we identify the last market to have the first trade, and acquire the most recent trade from the other markets, forming the first time tuple, and so on (called in their paper as the ‘replace all’ method). Table 3.8 has the initial and final number of observations given the aggregation process.

Table 3.1
Data set size and number of entries

Size for txt files extension and Microsoft SQL Server are showed in gigabytes, total number of entries is in million and daily average number of entries is in thousands.

	Size txt (GB)	Size SQL (GB)	Number of Entries (M)	Daily Avg Number of entries (K)
Trades	12.6	17.3	105	213

Figure 3.1
Intersection in trading hours (São Paulo time, UTC –3 hours)

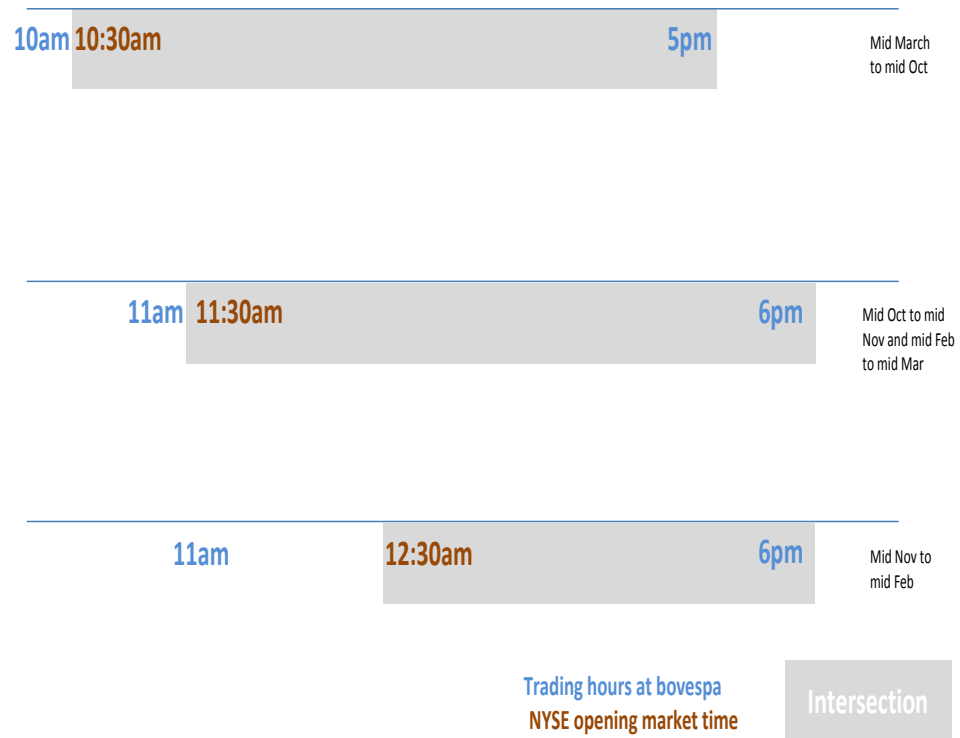


Table 3.2
2008 Most traded stocks at Bovespa

Stocks listed in the table are: Petrobras preferred shares (PETR4), Vale common (VALE5), Itau preferred (ITAU4), Bradesco preferred (BBDC4), Gerdau preferred (GGBR4), CSN common (CSNA3), Vale common different class (VALE3), Itausa preferred (ITSA4), Banco do Brasil common (BBAS3), Usiminas common (USIM5), Petrobras common (PETR3), Unibanco units (UBBR11), BVMF Bovespa common (BVMF3), Cemig preferred (CEMIG4), Sadia preferred (SDIA4), Americanas preferred (LAME4), ALL units (ALL11), Cyrela common (CYRE3), BMEF Bovespa common (BMEF3), TIM preferred (TCSL4). Bovespa accounts for all stocks traded in 2008, 20+ accounts for the sum of the stocks listed above.

Stocks	Total No of Trades (M)	Daily Avg (No trades) (K)	Total Volume (USD B)	Daily Avg (USD M)	Total quant (shares) (B)
Bovespa	44.7	179.4			
20+	24.4	98.1	442.0	1,775.0	27.4
PETR4	4.8	19.4	123.7	496.8	5.5
VALE5	3.5	14.0	85.5	343.3	4.0
ITAU4	1.3	5.3	20.2	80.9	1.1
BBDC4	1.3	5.3	22.9	91.9	1.2
GGBR4	1.3	5.2	16.5	66.5	0.9
CSNA3	1.2	4.8	21.2	85.0	0.7
VALE3	1.1	4.2	22.1	88.8	0.9
ITSA4	1.0	4.0	10.6	42.6	2.0
BBAS3	1.0	3.9	13.4	53.6	1.0
USIM5	1.0	3.8	21.5	86.4	0.6
PETR3	0.9	3.8	22.6	90.9	0.8
UBBR11	0.9	3.8	13.4	54.0	1.3
BVMF3	0.9	3.5	6.9	27.8	2.1
CMIG4	0.7	2.8	8.6	34.7	0.5
SDIA4	0.7	2.7	4.2	16.8	0.9
LAME4	0.7	2.6	5.0	19.9	0.8
ALL11	0.6	2.6	7.8	31.1	0.8
CYRE3	0.6	2.5	6.2	24.9	0.6
BMEF3	0.5	2.0	7.3	29.4	0.8
TCSL4	0.5	2.0	2.4	9.7	0.9

Table 3.3
2009 Most traded stocks at Bovespa

Stocks listed in the table are: Petrobras preferred shares (PETR4), Vale common (VALE5), BVMF Bovespa common (BVMF3), Gerdau preferred (GGBR4), Bradesco preferred (BBDC4), Itaúsa preferred (ITSA4), Usiminas common (USIM5), Vale common different class (VALE3), Petrobras common (PETR3), Itaú Unibanco preferred (ITUB4), CSN common (CSNA3), Banco do Brasil common (BBAS3), Cyrela common (CYRE3), Redecard common (RDCD3), Aracruz preferred (ARCZ6), Cemig preferred (CEMIG4), ALL units (ALL1), Gafisa common (GFSA3), Itaú Preferred (ITAU4), Americanas preferred (LAME4). Bovespa accounts for all stocks traded from January to December 2009, 20+ accounts for the sum of the stocks listed above.

Stocks	Total No of Trades (M)	Daily Avg (No trades) (K)	Total Volume (USD B)	Daily Avg (USD M)	Total quant (shares) (B)
Bovespa	58.0	256.4			
20+	28.9	128.0	329.7	1,458.7	29.4
PETR4	4.1	18.0	71.4	315.8	4.6
VALE5	4.0	17.5	67.6	298.9	4.1
BVMF3	2.2	9.7	19.2	84.8	3.7
GGBR4	1.9	8.4	15.2	67.4	1.5
BBDC4	1.6	7.0	17.4	77.0	1.2
ITSA4	1.5	6.7	9.1	40.3	1.9
USIM5	1.3	5.7	14.6	64.5	0.8
VALE3	1.2	5.3	17.1	75.9	0.9
PETR3	1.1	5.0	17.3	76.4	0.9
ITUB4	1.1	4.9	13.9	61.6	0.8
CSNA3	1.1	4.9	12.7	56.0	0.6
BBAS3	1.1	4.8	8.7	38.6	0.8
CYRE3	1.0	4.4	6.3	27.9	0.7
RDCD3	0.9	4.0	8.8	38.7	0.6
ARCZ6	0.9	3.9	5.1	22.5	3.3
CMIG4	0.9	3.8	5.5	24.4	0.4
ALLL1	0.9	3.8	5.3	23.2	0.9
GFSA3	0.8	3.6	4.8	21.4	0.5
ITAU4	0.7	3.3	6.8	30.1	0.6
LAME4	0.7	3.1	3.0	13.4	0.6

Table 3.4
Liquidity of Bovespa across NYSE opening times

Average percentage of overall trades in a given day in a 30-minute interval. 'B' refers to Bovespa trading hours and 'N' refers to NYSE trading hours.

Transaction Time: 1st Quarter 2008													
Time Interval													
from:		10:30	11:00	11:30	12:00	12:30	13:00	13:30					
to:		<10:00	11:00	11:30	12:00	12:30	13:00	13:30	14:00				
from 01 31/Jan	B:11-6,N:12:30-7	0%	0%	7%	7%	6%	9%	8%	7%				
from 01 to 16/Feb	B:11-6,N:12:30-7	0%	0%	6%	6%	5%	8%	9%	7%				
from 17 to 28/Feb	B:11-6,N:11:30-7	0%	0%	8%	10%	9%	8%	6%	5%				
from 1 to 09/mar	B:11-6,N:11:30-7	0%	0%	7%	9%	8%	7%	6%	5%				
from 10 to 31/mar	B:10-5,N:10:30-5	6%	9%	8%	7%	7%	6%	5%	5%				

Transaction Time: 2nd Quarter 2008													
Time Interval													
from:		10:30	11:00	11:30	12:00	12:30	13:00	13:30					
to:		<10:00	11:00	11:30	12:00	12:30	13:00	13:30	14:00				
Apr-08	B:10-5,N:10:30-5	6%	9%	8%	7%	6%	6%	6%	6%				
May-08	B:10-5,N:10:30-5	8%	10%	8%	8%	7%	6%	5%	5%				
Jun-08	B:10-5,N:10:30-5	6%	9%	8%	8%	7%	6%	5%	5%				
2Q08	B:10-5,N:10:30-5	6%	9%	8%	7%	7%	6%	5%	5%				

Table 3.5
Sample sizes before and after discarding outlier

We filter out any price entry p_i that does not conform to $|p_i - \bar{p}_i(k, 0.10)| < 3s_i(k) + 0.01$, where $\bar{p}_i(k, 0.10)$ and $s_i(k)$ are respectively the 10%-trimmed sample mean and the sample standard deviation of a neighborhood of k observations around i . We fix k according to the trade intensity, ranging from 20 to 60 observations. The column ‘trading platform’ informs the market at which the asset trades, ‘company’ reports whether the asset refers to Petrobras or Vale, ‘class’ reveals whether the share class is common (ON) or preferred (PN), and ‘symbol’ documents the asset symbol in the trading platform. We report the sample sizes (in millions) for both raw and clean data, i.e., respectively before and after excluding outliers.

trading platform	company	class	symbol	raw data	outliers	clean data
BM&FBovespa	Petrobras	ON	PETR3	2.11	4,812	2.10
		PN	PETR4	9.07	7,353	9.06
	Vale	ON	VALE3	2.07	8,139	2.06
		PN	VALE5	6.39	5,236	6.38
NYSE	Petrobras	ON	PBR.N	7.91	3,318	7.91
		PN	PBRa.N	5.02	4,485	5.02
	Vale	ON	RIO.N	6.93	1,159	6.93
		PN	RIOp.N	3.58	1,823	3.58
NYSE Arca	Petrobras	ON	PBR.P	11.82	3,460	11.82
		PN	PBRa.P	4.87	2,501	4.87
exchange rate	BRLUSD			4.09	600	4.09

Table 3.6
Data Cleaning Details

Number of observations (transaction prices) as raw observations in million, number of outliers in thousands, and final number of observations in million. PETR and PBR stands for Petrobras shares, VALE and RIO for Vale shares, AMBV and ABV for Ambev, BRTO and BTM for Brasil Telecom, GGBR and GGB for Gerdau and BBDC and BBD for Bradesco. We identify the outliers using the filter proposed by Brownlees and Gallo (2006).

		Stock	Raw obs (M)	outliers (K)	final obs (M)
		BRLUSD	4.088	0.600	4.087
Brazil	Pref	<i>PETR4</i>	9.071	7.353	9.063
	Com	<i>PETR3</i>	2.109	4.812	2.104
	Pref	<i>VALE5</i>	6.385	5.236	6.380
	Com	<i>VALE3</i>	2.066	8.139	2.058
	Pref	<i>AMBV4</i>	0.720	4.109	0.716
	Pref	<i>BRTO4</i>	0.555	1.564	0.554
	Pref	<i>GGBR4</i>	3.237	3.000	3.234
	Pref	<i>BBDC4</i>	2.958	3.909	2.954
Nyse	Pref	<i>PBRa.N</i>	5.021	4.485	5.017
	Com	<i>PBR.N</i>	7.914	3.318	7.910
	Pref	<i>RIOp.N</i>	3.577	1.823	3.575
	Com	<i>RIO.N</i>	6.930	1.159	6.929
	Pref	<i>ABV_N</i>	1.119	1.645	1.118
	Pref	<i>BTM_N</i>	0.200	0.521	0.200
	Pref	<i>GGB_N</i>	2.829	0.723	2.828
	Pref	<i>BBD_N</i>	3.599	0.959	3.598
Arca	Pref	<i>PBRa.P</i>	4.873	2.501	4.870
	Com	<i>PBR.P</i>	11.821	3.460	11.818
	Pref	<i>ABV_P</i>	0.506	1.190	0.505
	Pref	<i>BTM_P</i>	0.106	0.391	0.106
	Pref	<i>GGB_P</i>	4.164	0.871	4.163
	Pref	<i>BBD_P</i>	6.091	1.038	6.090
Total			89.939	62.806	89.876

Table 3.7
Aggregating tick data into fixed time intervals

We aggregate tick data into regular intervals of 30 seconds, giving way to a sample size of over 350,000 observations for each market price from January 2008 to November 2009. The first 4 columns are as in Table 3.5. The column ‘tick data’ reports the initial sample size of irregularly-spaced-in-time data (in millions). It differs from the sample size of the clean data in Table 3.5 mainly because of holidays in Brazil and in the US. Finally, ‘missing’ informs how many 30-second intervals feature at least one missing observation across the different markets and share classes (also in millions), whereas ‘zero returns’ documents the proportion of zero returns due to missing observations.

trading platform	company	class	symbol	tick data	missing	zero returns
BM&FBovespa	Petrobras	ON	PETR3	1.94	0.10	28%
		PN	PETR4	7.79	0.00	1%
	Vale	ON	VALE3	2.06	0.09	25%
		PN	VALE5	6.38	0.01	2%
NYSE	Petrobras	ON	PBR.N	7.42	0.00	1%
		PN	PBRa.N	4.80	0.02	5%
	Vale	ON	RIO.N	6.92	0.01	2%
		PN	RIOp.N	3.57	0.04	10%
NYSE Arca	Petrobras	ON	PBR.P	11.19	0.01	3%
		PN	PBRa.P	4.63	0.04	13%
exchange rate	BRLUSD			2.88	0.03	10%

Table 3.8
Data Aggregation

Number of initial observations in million, number of missing observations and aggregated observations in thousands. PETR and PBR stands for Petrobras shares (3: common, 4 and a: preferred, N:NYSE, P: ARCA), VALE and RIO for Vale shares (3: common, 5 and p: preferred, N:NYSE), AMBV and ABV for Ambev (all preferred, N:NYSE and P:ARCA), BRTO and BTM for Brasil Telecom (all preferred, N:NYSE and P:ARCA), GGBR and GGB for Gerdau (all preferred, N:NYSE and P:ARCA), BBDC and BBD for Bradesco (all preferred, N:NYSE and P:ARCA) and BRLUSD for the Brazilian Reais/US dollar exchange rate. We aggregate the data into time tuples using the methodology proposed by Harris, McInish, Shoesmith, and Wood (1995).

Freq.	Stock	Initial obs (M)	Missing obs (k)	Agg. Obs (k)	% Missing
30"	<i>PETR4</i>	7.79	4.53	352.68	1%
	<i>PETR3</i>	1.94	100.07	352.68	28%
	<i>PBRa_P</i>	4.63	44.20	352.68	13%
	<i>PBR_P</i>	11.19	9.95	352.68	3%
	<i>PBRa_N</i>	4.80	16.66	352.68	5%
	<i>PBR_N</i>	7.42	4.86	352.68	1%
	<i>BRLUSD</i>	2.84	34.47	352.35	10%
	<i>VALE5</i>	6.38	6.28	352.35	2%
	<i>VALE3</i>	2.06	88.46	352.35	25%
	<i>RIO_N</i>	6.92	8.76	352.35	2%
	<i>RIOp_N</i>	3.57	36.89	352.35	10%
	<i>BBDC4</i>	2.60	123.32	352.18	35%
	<i>BBD_N</i>	3.21	29.21	352.18	8%
	<i>BBD_P</i>	5.53	68.91	352.18	20%
	<i>GGBR4</i>	2.78	123.88	352.16	35%
	<i>GGB_N</i>	2.55	38.53	352.16	11%
	<i>GGB_P</i>	3.78	96.73	352.16	27%
90"	<i>BRLUSD</i>	2.84	5.72	117.49	5%
	<i>AMBV4</i>	0.65	21.19	117.49	18%
	<i>ABV_N</i>	1.01	10.41	117.49	9%
	<i>ABV_P</i>	0.46	31.39	117.49	27%
300"	<i>BRLUSD</i>	2.84	0.89	35.23	3%
	<i>BRTO4</i>	0.49	1.63	35.23	5%
	<i>BTM_N</i>	0.18	5.97	35.23	17%
	<i>BTM_P</i>	0.10	12.70	35.23	36%

Figure 3.2
Price Evolution

Displays price evolution of Ambev, Gerdau, Bradesco, BR Telecom, Petrobras and Vale stocks traded at BOVESPA, NYSE and ARCA. Sampling frequency are fixes as follows: Ambev, 90"; Gerdau, 60"; Bradesco, 30"; Petrobras, 300"; and Vale 30". Prices are aggregated following Harris, McNish, Shoemith, and Wood (1995) and free of non plausible values.

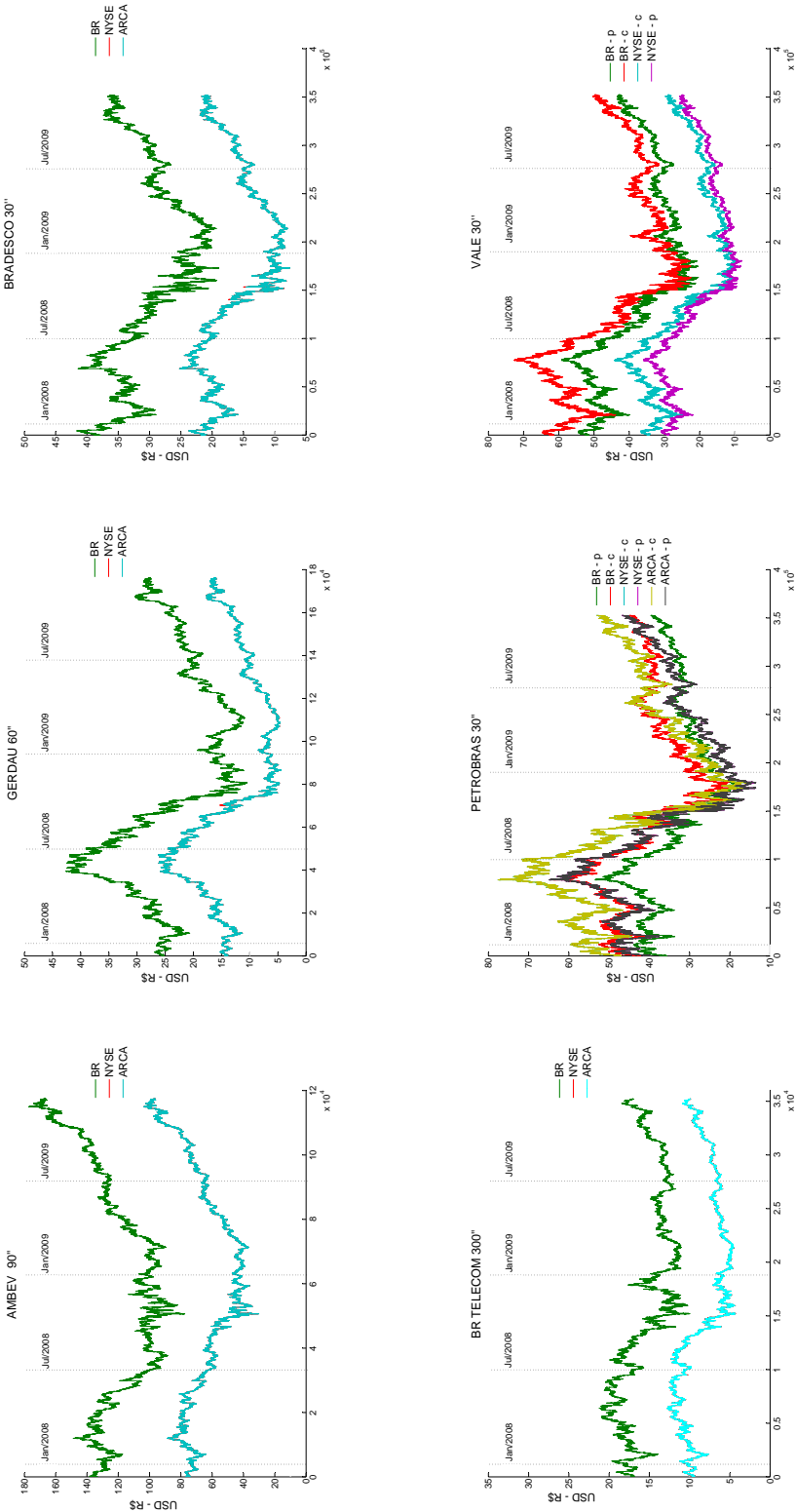


Figure 3.3

The prices of Petrobras' shares and their ADR counterparts

The first plot depicts share prices in Brazilian reais and ADR prices in US dollars, whereas the second chart displays all prices in US dollars. PETR3 and PETR4 correspond to common and preferred shares at the BM&FBovespa, respectively. Similarly, PBR and PBRa are the symbols for Petrobras' common and preferred ADRs, with extensions indicating the trading platform: N for NYSE and P for Arca.

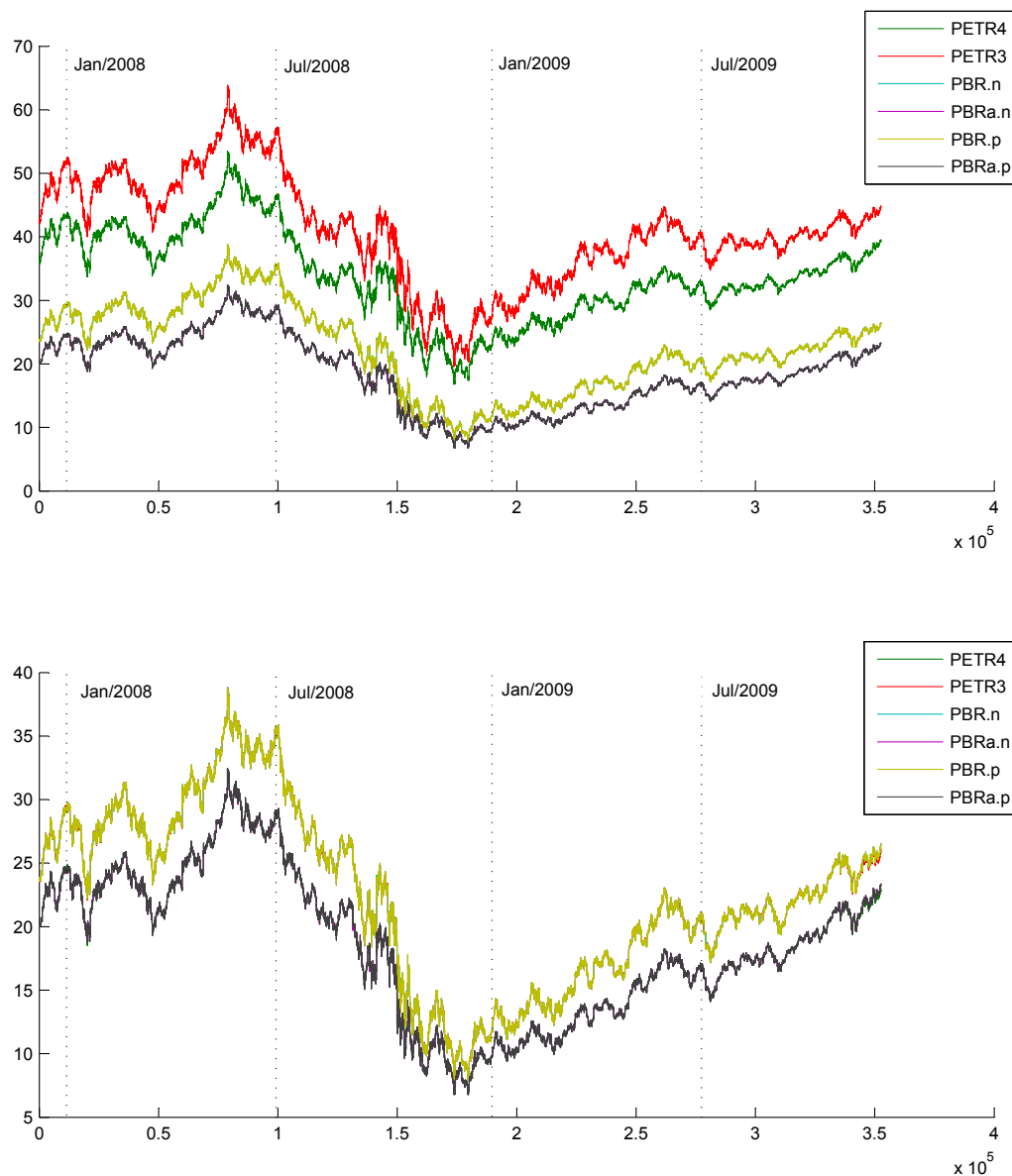
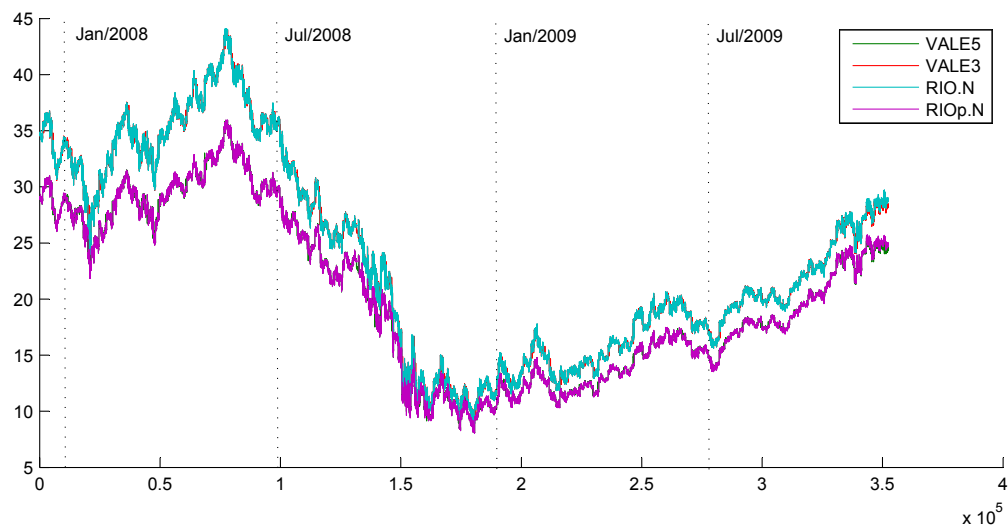
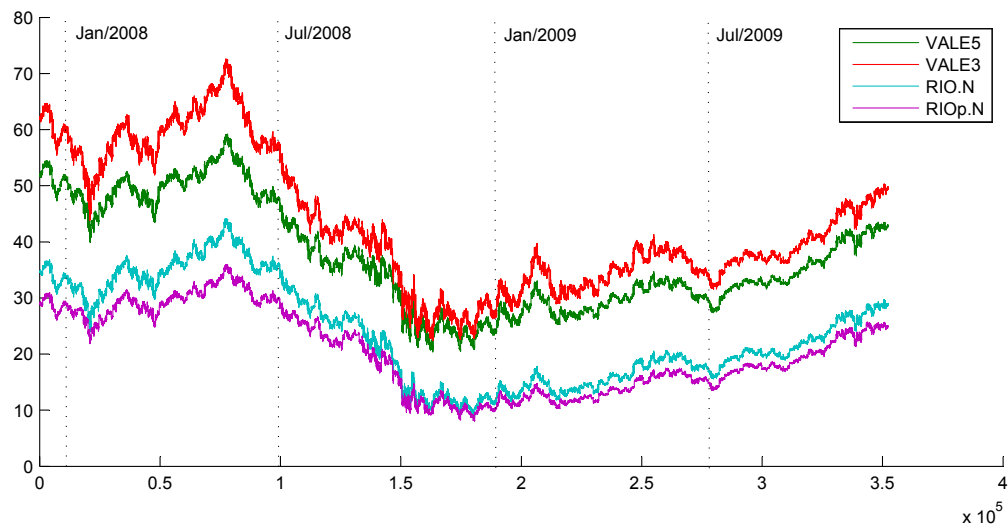


Figure 3.4

The prices of Vale's shares and their ADR counterparts

The first plot depicts share prices in Brazilian reais and ADR prices in US dollars, whereas the second chart displays all prices in US dollars. VALE3 and VALE5 correspond to common and preferred shares at the BM&FBovespa, respectively. Similarly, RIO and RIOp are the symbols for Vale's common and preferred ADRs at the NYSE.



Chapter 4

An order invariant Measure for Price Discovery

4.1 Introduction

Price discovery has recently become a hot topic mainly for two reasons. First, the increasing availability of high-frequency data allows studying how efficiently and timely each market reacts to news in a much more precise manner. Second, quantitative trading strategies that rely on price discovery analyses (e.g., pair trading) are nowadays responsible for a substantial amount of assets under management. This chapter extends the standard price discovery methodology to deal with dual-class assets traded on multiple markets. The idea is to exploit every piece of information we have about the fundamental value of a firm by looking at the prices of both common and preferred shares across different trading platforms. As a by-product, by looking at the difference between the prices of the common and preferred shares, we may also shed some light on the behavior of the

dual-class premium.

The main technical difficulty is to contrive a unique price discovery measure that does not assume *a priori* which share class and/or market lead the impounding of new information. For the standard information share (IS) measure of price discovery Hasbrouck (1995), which gauges the fraction of the variance of the fundamental price innovation due to the variance of a given asset/market price innovation, one normally imposes a triangular structure from the most informative to least informative market price in order to handle contemporaneous correlation. The information share of the price of a given share class at a given trading platform thus depends on the specific ordering we employ. This is definitely a problem if one wishes to keep agnostic about lead-lag patterns.

There are two standard solutions for the non-uniqueness of the IS measure in the literature. The first is to increase the sampling frequency at which we record prices hoping for less contemporaneous correlation between the price innovations. The idea is that one-way causality at the high frequency could well dissolve into contemporaneous correlation at the low frequency. However, there is unfortunately no guarantee that this works in practice, especially for dual-class shares. The second is to consider the average IS across different orderings of market prices. This is a simple and, most likely, effective solution if there are only a few market prices. However, as the number of assets/markets increase, one would have to average over thousands of information shares as there are a factorial number of possible orderings. For instance, a system consisting of 7 market prices as in Section 4.3 would lead to the unreasonable amount of $7! = 5,040$ distinct orderings.

The first contribution of this chapter is methodological. We derive a variant of the IS measure that rests on the spectral decomposition of the covariance matrix of the price innovations. The latter decomposition is unique and order invariant. As a result, our measure of information share is completely agnostic about which market price reacts first to new information. This is especially important for the case of dual-class shares because we have no reason to believe that one share class (or market) is relatively more informative than the others. In addition, Monte Carlo simulations show that the spectral-based IS measure works pretty well in finite samples as opposed to the standard measure in the presence of contemporaneous correlation.

Methodologically, the closest paper to ours is Lien and Shrestha (2009). They also propose an order-invariant IS measure. It is quite similar to ours, though much more complicated for it involves a decomposition of the correlation matrix rather than of the covariance matrix. It is not clear what is the economic intuition behind their more complicated method, moreover it does not yields better performance results. Grammig and Peter (2012) achieve unique identification for the IS measure by imposing tail dependence restrictions. Their identification strategy is very ingenious, relying on the distinctive market microstructure of each trading platform. However, it requires the econometrician to take a stand on how the shocks disseminate across markets. In contrast, our spectral-based procedure is completely agnostic, keeping the reduced-form philosophy of the original IS measure.

Our contribution is not only methodological, though. We also empirically investigate price discovery in dual-class shares trading both at the

Sao Paulo Stock Exchange (BM&FBovespa) and at the New York Stock Exchange (NYSE) through the American Depositary Receipt (ADR) program. This means investigating price discovery using a much richer data set than previous studies. It is richer because it takes advantage of the fact that, dual-class premium aside, both common and preferred stock prices depend on the latent efficient/fundamental stock price. The focus on Brazilian stocks and their ADRs is convenient for a number of reasons. First, the BM&FBovespa is the leading exchange in Latin America and among the 10 largest stock exchanges in the world. Second, the trading hours at the BM&FBovespa track to a large extent the trading hours at the NYSE, amounting to an overlapping of 6.5 hours from mid-February to mid-November and of 5.5 hours in the remaining 3 months of the year. This comes as a huge advantage relative to most studies in price discovery, which end up with only 2 to 3 hours of intersection for using European stocks and their ADR counterparts. Third, preferred shares are historically very liquid in the BM&FBovespa because Brazilian firms could issue two preferred shares for each common share before 2001 (now it is a one-to-one ratio). The number of common shares over the number of preferred shares is indeed about 0.75 for Petrobras and 0.65 for Vale. Fourth, quality transactions data from the BM&FBovespa are available from December 2007 to November 2009, allowing us to examine how price discovery works over different market cycles.

We restrict attention to the two most liquid stocks in Brazil, namely, Petrobras and Vale, whose common and preferred shares also trade as ADRs at the NYSE. Note that, for Petrobras, we also able to employ ADR trades and quotes from Arca (previously known as Archipelago Exchange

or ArcaEx), NYSE's Chicago-based electronic platform. The latter is the second largest electronic communication network in the world, accounting for roughly 10% of NYSE-listed securities traded and 20% of Nasdaq-listed securities traded. This amounts to a system of 7 variables: common and preferred share prices in the BM&FBovespa, Arca and NYSE, plus the exchange rate. We include the latter so as to gauge how stock prices adjust to exchange-rate shocks.

Our price discovery analysis yields some interesting findings. First, the US market is at least as informative as the home market for both Petrobras and Vale. This is not so surprising given that these Brazilian behemoths are commodity exporters and hence more sensitive to international (rather than local) market conditions. Second, we evince that Petrobras' common shares are more informative than preferreds in the US and vice-versa in Brazil. This seems to derive from liquidity issues given that the trade intensity is higher exactly for these class-market combinations. In contrast, common and preferred shares have a similar role in Vale's price discovery process. This illustrates the fact that Vale's common shares may actually entail control power, as opposed to the case of the state-owned Petrobras. Third, we find that the exchange rate seems to react to changes in the efficient prices of Petrobras and Vale (possibly due to the omission of commodity indices in the analysis). Fourth, shocks in the dual-class premium entail a permanent impact in normal times, whereas their effects are transitory during the financial crisis. We argue that the latter is consistent with a dual-class premium as a function of private benefits that shareholders may obtain for holding voting rights (see Zingales 1994, 1995). As there are fewer opportunities to extract private benefits, investors cease to price the

dual-class premium as an asset in periods of financial distress. Up to our knowledge, this is the first work to provide evidence that the price discovery mechanism may change across market cycles.

The remainder of this chapter is as follows. Section 4.2 develops the spectral-based information share measure that is more suitable to study price discovery in large price systems. Section 4.3 documents the empirical price discovery analyses for Petrobras and Vale. We relegate Section 4.4 a Monte Carlo study of the performance of the spectral-based IS measure relative to the extant IS measures in the literature. Section 4.5 offers some concluding remarks.

4.2 Information share in a large price system

To allow for common and preferred shares in both domestic and foreign markets, we first extend the three-variable model proposed by Grammig, Melvin, and Schlag (2005) and then modify Hasbrouck's (1995) IS methodology so as to ensure uniqueness of the price discovery measure. The setup is such that every stock price in the system shares a common component given by the fundamental value of the firm (i.e., the present value of the firm's expected cash flow). This means that these prices cointegrate in that they should not diverge too much from each other because they must track somehow the implicit efficient price. However, the latter is not the only common factor driving the system dynamics. To make stock prices in the foreign market comparable to stock prices in the domestic market, one must include the exchange rate in the system as in Grammig, Melvin, and

Schlag (2005). This results in another common factor, which relates to the efficient exchange rate. Note that the latter may differ from the observed exchange rate due to transitory market microstructure effects.

In our setup, the dual-class premium stands for another potential common factor. In that case, the gap between common and preferred share prices gauges the dual-class premium, up to transient effects (e.g., liquidity issues). In principle, the dual-class premium stands for the price of voting rights (see Zingales 1994, 1995). It thus relates to the fundamental value of the firm through at least three channels (as in Chapter 5). First, it depends on whether the investor is able to extract private benefits from holding voting rights. Such opportunities are more likely in boom periods, when the value of the firm is higher. Second, it also reflects the expected takeover premium paid to shareholders outside the control block. This implies a premium that increases with voting power, but decreases with ownership, size and trading liquidity Smith and Amoako-Adu (1995). Finally, the third channel is through a principal-agent problem. Stronger voting rights induce better monitoring of the board of directors. As such, positive shocks to the dual-class premium may reduce principal-agent concerns, increasing the value of the firm.

Regardless of the number of common factors governing the price dynamics, it remains the fact that common and preferred share prices must not drift apart, otherwise arbitrage opportunities would persist. There are several ways to represent such a cointegrated system. For instance, the vector error correction model (VECM) posits that

$$\Delta y_t = \xi_0 y_{t-1} + \xi_1 \Delta y_{t-1} + \xi_2 \Delta y_{t-2} + \dots + \xi_p \Delta y_{t-p} + \zeta + \epsilon_t,$$

where $\xi_0 = \alpha\beta'$, α is the error correction term, β is the cointegrating vector, and y_t is a vector of prices for both share classes and markets (including the exchange rate). We further assume that ϵ_t is a zero-mean white noise with a covariance matrix given by Ω and that ζ is such that cumulative price changes feature no deterministic time trends.

Albeit the VECM representation is amenable to estimation as well as to economic interpretation, it is not unique. There are actually infinitely many error-correction representations, though they all lead to the same vector moving average (VMA) representation:

$$\Delta y_t = \epsilon_t + \psi_1 \epsilon_{t-1} + \psi_2 \epsilon_{t-2} + \dots = \Psi(L) \epsilon_t.$$

Hasbrouck (1995) thus propose to recover the VMA coefficients from the VECM estimates and then apply a Beveridge-Nelson random-walk decomposition. This results in $\psi \epsilon_t$ as the vector of common factor innovations, with ψ denoting a non-square matrix that discards any repeated row of the moving-average impact matrix $\Psi(1)$. The covariance matrix of the innovation vector then is $\psi\Omega\psi'$. If the latter is diagonal, Hasbrouck (1995) defines the information share as the relative contributions of each share class/market to the total variation of the innovation in the permanent common factor.

However, the covariance matrix Ω of the reduced-form errors is no longer diagonal in the presence of contemporaneous correlation between markets, invalidating the above procedure. To circumvent this, Hasbrouck (1995) proposes the use of a Cholesky decomposition of Ω . This amounts to assuming a lower-triangular structure in the system, with market prices sorted from least to most endogenous. As a result, the IS measure is not unique,

varying with the ordering of the prices. This is particularly inconvenient in the context of dual-class shares in multiple markets. The number of possible permutations increases at a factorial rate with the system dimension. A stock with both common and preferred shares trading at the domestic and foreign markets would compose a system with (at least) 5 price series, implying over 1,000 different orderings. This is likely to entail a large gap between the minimum and maximum information shares, impairing any sort of meaningful price discovery analysis. Huang (2002), Hupperets and Menkveld (2002), Kim (2010b), Kim (2010a), and Grammig and Peter (2012) indeed report sizeable differences even for systems of only two/three market prices.

To derive an order-invariant IS measure, we employ a spectral decomposition of Ω . The resulting IS measure is the ratio of $[\psi S]_{ij}^2$ to $[\psi \Omega \psi']_{ii}$, where $S = \Omega^{1/2} = V \Lambda^{1/2} V'$, with Λ and V respectively denoting the diagonal matrix with the eigenvalues along the principal diagonal and the matrix with the corresponding eigenvalues in the columns. In stark contrast with the Cholesky factorization, the spectral decomposition is completely agnostic about lead-lag patterns, imposing no assumption about which share class or market is more informative. This makes our framework particularly suitable to identify which markets are dominant in setting the price Garbade and Silber (1983).

Our spectral-based IS measure of contribution to the price discovery is very similar in spirit to Lien and Shrestha's (2009). In particular, they suggest an alternative IS measure that rests on the spectral decomposition of the correlation matrix (rather than of the covariance matrix). This brings about unnecessary complications because one must back out the implicit

decomposition of the covariance matrix from the spectral factorization of the correlation matrix to compute the information share. Monte Carlo simulations in Section 4.4 indeed show that it pays off to take a more direct approach based on the eigendecomposition of the covariance matrix.

4.3 Which share class leads, and in which market?

We expect the dynamics of share and ADR prices to feature no more than three common factors. The first corresponds to the efficient exchange rate in view that the system must include the BRL/USD exchange rate to make ADR prices in US dollars comparable to share prices at the BM&FBovespa. The second refers to the fundamental values of Petrobras and Vale given by the present value of their expected cash flow. Note that CVM normally requires preferred shares to pay 10% more of preferential dividends relative to common shares (as calculated from a *minimum* dividend payment of 25% of the adjusted net income). However, both Petrobras and Vale distribute systematically more dividends than the minimum payment that CVM requires. As such, their common and preferred shares end up receiving the same amount of dividends and hence the same present value of expected cash flow.

The dual-class premium may stand as a third stochastic trend in the system. Note that the Brazilian government detains the vast majority of Petrobras voting shares and hence it makes no sense to speak about takeover premium. As the private benefits story, it seems to fit the bill for both Petrobras and Vale. The Brazilian government has been imposing a

gasoline price cap on Petrobras since 2006 to help control inflation (see, e.g., *The Economist*, “The perils of Petrobras: How Graça Foster plans to get Brazil’s oil giant back on track”, November 17, 2012). Surprisingly, the same arguments also apply to Vale. Although it has been privatized in 1997, the Brazilian government indirectly detains the majority of the voting rights through a consortium of state pension funds. This not only makes takeovers very unlikely, but also raises the issue that the government may exert sway on Vale against the interest of the minority shareholders. For instance, the former CEO of Vale, Roger Agnelli, was ousted in 2011 by the state pension funds because he did not invest enough at home, particularly in low-margin industries such as steel and shipbuilding (see, e.g., *The Economist*, “Vale dumps its boss: Roger and out”, April 1st, 2011).

Given their generous dividend policy, one would expect preferred shares to be more appealing to investors than common shares, therefore commanding a premium, in the absence of takeover risk. That is not the case, though. Their common share prices are superior to their preferred prices in Brazil and in the US. Further, liquidity premium does not suffice to justify the dual-class premium, otherwise the sign of the latter in Brazil would differ from the sign in the US. Indeed, preferred shares are much more actively traded at the BM&FBovespa than common shares for both Petrobras and Vale, whereas the opposite is true for their ADR counterparts. The fact that the difference between the common and preferred share prices is positive regardless of the trading platform perhaps indicates that the foreign market leads the process of impounding information for Petrobras and Vale. We thus conjecture that their common ADRs should play a major role in

the price discovery mechanism.

In what follows, we first describe the results for Petrobras and then discuss the findings for Vale. Note that the main difference between the two analyses is that we only observe prices at the NYSE Arca for Petrobras. The price system for Vale thus consists of the prices of common and preferred shares at the BM&FBovespa as well as their American Deposit shares at the NYSE (i.e., 5 variables, including the exchange rate), whereas the Petrobras system also includes the ADR counterparts at the Arca trading platform. Note that we actually expect Arca to impound information more timely for Petrobras than the NYSE. Arca's smart order router does not restrict attention exclusively to NYSE's quotes, executing orders at the trading venue with the best available quote across all stock exchanges in the US (including NASDAQ). Finally, to better understand how the price discovery mechanism changes across different market cycles, we estimate IS measures for the periods ranging from January to June 2008, July to December 2008, January to June 2009, and July to November 2009.

4.3.1 Petrobras

Figure 3.3 plots the prices of the Petrobras shares at the BM&FBovespa (in both BRL and USD terms) as well as their corresponding ADR prices in the US market. It is striking how the prices move in tandem, even if not surprising, given that they all relate to the same fundamental value. We separate the subperiods we consider by dashed lines so as to highlight how different they are. Petrobras share prices are clearly trending up in the first subsample running from January to June 2008, but then stock prices plummet in the second half of 2008 as a reaction to the steady decline

in the price of oil. Petrobras share prices show some recovery in the last two subsamples, reflecting to some extent the steady rise in oil prices as from January 2009. Share prices do not recover fully probably because of investors' fears that Petrobras' primary *raison d'être* is to serve the nation in whatever way the Brazilian government sees fit rather than to make a profit.

For each subperiod, we carry out a price discovery analysis relying on the spectral-based IS measure of Section 4.2. We bootstrap the VECM residuals as in Li and Maddala (1997) to compute the standard errors of the information share. In particular, we consider 1,000 bootstrap samples. The top panel of Table 4.1 reports the results for the first half of 2008. There are 4 cointegrating vectors and hence 3 common factors. The first cointegrating vector takes the difference between NYSE and Arca prices of Petrobras common shares. As both these ADRs have voting rights and prices in US dollars, their price difference essentially eliminates the common factor given by the fundamental value of Petrobras. Voting rights aside, the same reasoning applies to the second cointegrating vector, which considers the difference between the prices of the preferred ADRs at the NYSE and Arca trading platforms.

The third cointegrating vector dictates that prices in Brazil and in the US must not drift apart once we consider them in the same currency. This indicates that the second common factor is attributable to the efficient exchange rate. Finally, the fourth cointegrating vector corresponds to the difference between the BM&FBovespa and NYSE *observed* dual-class premia.¹ This means that the dual-premium class indeed is a common factor

¹ The price gap between common and preferred shares differs from the latent dual-class premium because of transient market microstructure effects.

driving the price dynamics, otherwise we would not have to take the difference between the observed dual-class premia in Brazil and in the US to get stationarity. This may come as a surprise, especially at such a high frequency. However, it is consistent with the Brazilian government expropriating preferred shareholders as a class during this period.

As for the IS estimates, the preferred share is much more informative than the common share in Brazil, whereas the opposite is true in the US. This may sound puzzling, but it actually reflects well the difference in their liquidity as seen in Section 3.3. The trading of common shares through the ADR program indeed responds for 20% of the total shares, which is extremely high in view that the Brazilian government detains about 55% of the common shares. Table 4.1 also confirms our prediction that Arca's smart order route contributes more to the impounding of information into security prices than the NYSE. Further, we also find that the exchange rate is not completely exogenous as one would normally expect Grammig, Melvin, and Schlag (2005). This is probably due to the fact that the system does not account for international oil prices, which affect both Petrobras share prices and the strength of the US dollar. In fact, the correlation between changes in the oil price and in the BRL/USD exchange rate is over 0.42 in the sample period. Finally, it is also interesting to observe that it is the US market that absorbs shocks in the efficient exchange rate.

The bottom panel of Table 4.1 reports the estimates of the spectral IS measures as well as of the cointegrating vectors for the second half of 2008. This is when the financial crisis finally hits Brazil: The IBOVESPA drops about one third of its value and the Brazilian real devaluates over 50% against the US dollar in this period. The financial distress seems to

strongly affect the price discovery process. To begin with, there are now 5 cointegrating vectors and hence only two common factors. In particular, the dual-class premium becomes stationary, characterizing the fifth cointegrating vector. The fact that investors do not price the voting premium anymore as an asset is still consistent with our private benefit story. It is much easier to expropriate the shareholders with no control power in periods of boom. As crises shut down most opportunities for extracting private benefits, the difference between common and preferred shares starts to reflect much more liquidity issues than anything else. Additionally, the contribution of Petrobras shares at the BM&FBovespa to the price discovery mechanism sinks in this period. This drop is particularly strong for the preferred shares. At the same time, the Arca platform gains in importance. As opposed to the first half of 2008, it is now the Brazilian market that incorporates shocks in the efficient exchange rate.

Table 4.2 documents a similar pattern for the first half of 2009 in that the dual-class premium remains stationary and the BM&FBovespa keeps losing importance in the price discovery process. In turn, the second half of 2009 resembles more the pre-crisis period, with the efficient exchange rate, the fundamental value of the company and the dual-class premium driving the stochastic trends in the system. The only difference is that the BM&FBovespa does not recover relative importance, whereas the NYSE starts playing a more significant role probably due to the increase in the frequency and value of block trades as from September 2009. This is when Brazil obtains the investment grade rating from Moody's, allowing foreign pension funds to invest in Brazilian ADRs.

As a robustness check, we estimate the IS measures using prices at the

60-second interval. The results are very similar and qualitatively exactly the same. We also carry out the price discovery analysis using the complete sample period (i.e., January 2008 to November 2009) as well as by years (i.e., January to December 2008 and January to November 2009). We find similar information share, confirming that the foreign market contributes more to the price discovery mechanism than the home market. This is particularly true for the ADR prices of the common shares and for Arca, ratifying that liquidity matters.

4.3.2 Vale

In the absence of enough trades at the Arca platform, we focus on a system of 5 market prices: common and preferred shares at the BM&FBovespa (VALE3 and VALE5, respectively) and their corresponding ADRs at the NYSE (RIO.N and RIOp.N, respectively), plus the exchange rate. We expect Vale to feature a price discovery process similar to Petrobras. As before, the system does not include international metal prices and hence we do not expect the exchange rate to move in a completely exogenous manner relative to Vale’s fundamental value. Note that the extension 5 in Vale’s preferred shares defines them as ‘class A’, so that preferred shareholders have the right to vote in General Assembly deliberations, just as common shareholders. The only difference is that preferred shareholders do not have a say in the composition of the Board of Directors. We thus expect Vale’s preferred shares to contribute relatively more to the price discovery than Petrobras’ preferreds.

Figure 3.4 displays the prices of Vale’s common and preferred shares and of their ADRs. The pattern it depicts is very similar to that of Petro-

bras in that the second half of 2008 witnesses a huge drop in prices, with a slow recovery afterwards. Table 4.3 reveals the information shares we obtain for each half of 2008 and 2009, respectively. As in the case of Petrobras, the dual-class premium is a common factor in the first half of 2008, but then becomes stationary from July 2008 to June 2009. In this turbulent period, the preferred shares lose most of their importance (especially in the NYSE) and hence the price discovery takes place through the common shares. Shocks in the dual-premium class regain its permanent impact only after July 2009. As before, the NYSE is more informative than the BM&FBovespa regardless of the share class. The contribution of the NYSE to the price discovery actually increases. Further, we also reject the exogeneity of the exchange rate. This is not surprising given that the sample correlation between the changes in the BRL/USD and in the S&P industrial metals spot index is pretty high at 0.53. We also find that it is the ADR prices that adjust for shocks in the efficient exchange rate.

The main difference relative to what we observe for Petrobras is that preferred shares play a much more significant part for Vale. The higher information share we uncover for the preferred ADRs are likely due to the ‘class A’ nature of VALE5. In contrast, common and preferred shares at the B&FBovespa entail similar contributions to the price discovery (though weaker than their ADR counterparts). The financial crisis seems to have a significant impact in this pattern. The information shares of the preferred shares are indeed much lower from July to December 2008, though they start to recover in the first half of 2009, regaining their full importance in the price discovery mechanism only by the second half of 2009. Also, we observe that, similarly to what happens with Petrobras, the BM&FBovespa

loses importance for the price discovery in Vale shares after the financial crisis. Finally, the second half of 2009 marks the return of the dual-class premium as a common factor driving the price dynamics.

Apart from sampling the prices at 60-second frequency, we also compute information shares for each year and for the overall sample. As before, we do not observe any qualitative change in the IS estimates. All in all, we conclude that (1) the foreign market impounds more information than the home market, (2) common and preferred shares have similar contributions to the impounding of information into securities prices, (3) the exchange rate is not entirely exogenous to the variations in Vale share prices, and (4) Vale's dual-class premium is a common factor only in normal times.

4.4 Simulations

This section examines the implications of decomposing the covariance matrix by Cholesky and by the spectral approach as what concerns the estimation of information share. We simulate from three different structural market microstructure models. In the simplest setup M_1 , the ADR price p_t^f follows the share price p_t^h at the home market and the exchange rate e_t is entirely exogenous, namely,

$$\begin{aligned} e_t &= e_{t-1} + u_t^e \\ p_t^h &= p_{t-1}^h + u_t^h \\ p_t^f &= p_{t-1}^h + e_{t-1} + u_t^f, \end{aligned}$$

where all prices are in logs and (u_t^e, u_t^h, u_t^f) is a vector of Gaussian white noises. In the other two settings, we also consider that there are both

common and preferred shares (indexed by subscripts c and p , respectively) at the home and foreign markets.

The model M_2 assumes that the prices of the common and preferred shares are independent at the home market and that the ADR prices in the foreign market follow their counterparts in the home market:

$$\begin{aligned} e_t &= e_{t-1} + u_t^e \\ p_{p,t}^h &= p_{p,t-1}^h + u_{p,t}^h \\ p_{c,t}^h &= p_{c,t-1}^h + u_{c,t}^h \\ p_{p,t}^f &= p_{p,t-1}^h + e_{t-1} + u_{p,t}^f \\ p_{c,t}^f &= p_{c,t-1}^h + e_{t-1} + u_{c,t}^f, \end{aligned}$$

where $(u_t^e, u_{c,t}^h, u_{p,t}^h, u_{c,t}^f, u_{p,t}^f)$ is a vector of Gaussian white noises. Last but not least, M_3 posits that the prices of the common share at the home market and of both ADRs in the foreign market follow the price of the preferred share at home market, that is to say,

$$\begin{aligned} e_t &= e_{t-1} + u_t^e \\ p_{p,t}^h &= p_{p,t-1}^h + u_{p,t}^h \\ p_{c,t}^h &= p_{p,t-1}^h + d + u_{c,t}^h \\ p_{p,t}^f &= p_{p,t-1}^h + e_{t-1} + u_{p,t}^f \\ p_{c,t}^f &= p_{p,t-1}^h + d + e_{t-1} + u_{c,t}^f, \end{aligned}$$

Note that both M_2 and M_3 assume a constant dual-premium class of d for the sake of simplicity.

We simulate 1,000 replications of every model, each with a sample size

of 10,000 observations. Note that we discard the first 500 observations in order to alleviate any dependence on the initial values. We consider two cases for the covariance matrix of the errors. The first imposes an identity covariance matrix, implying a unique Cholesky decomposition that does not vary with the ordering of the variables. The second case assumes the following nondiagonal covariance matrices:

$$\Omega_1 = \begin{pmatrix} 1 & 0.4 & 0.1 \\ 0.4 & 1 & 0.5 \\ 0.1 & 0.5 & 1 \end{pmatrix}, \quad \Omega_2 = \begin{pmatrix} 1 & 0.1 & 0.3 & 0.4 & 0.1 \\ 0.1 & 1 & 0.2 & 0.4 & 0.4 \\ 0.3 & 0.2 & 1 & 0.4 & 0.4 \\ 0.4 & 0.4 & 0.4 & 1 & 0.2 \\ 0.1 & 0.4 & 0.4 & 0.2 & 1 \end{pmatrix} \quad \text{and}$$

$$\Omega_3 = \begin{pmatrix} 1 & 0.5 & 0.5 & 0.2 & 0.5 \\ 0.5 & 1 & 0.7 & 0.8 & 0.9 \\ 0.5 & 0.7 & 1 & 0.5 & 0.7 \\ 0.2 & 0.8 & 0.5 & 1 & 0.7 \\ 0.5 & 0.9 & 0.7 & 0.7 & 1 \end{pmatrix},$$

where Ω_j is the covariance matrix for the model M_j . The idea is to assess the behavior of the IS measures based on the Cholesky decomposition in view that the ordering of the variables now matters.

Table 4.4 documents the true information share and their estimates based on the Cholesky and spectral decompositions for the case of diagonal covariance matrix. For the sake of brevity, we report the results only for M_1 because both estimators perform extremely well regardless of the setup we consider. In particular, they are both very accurate and precise, featuring no bias in the IS estimation. This means that the price we pay for the

agnosticism of the eigendecomposition is negligible.

Tables 4.5 to 4.7 report the results for the nondiagonal covariance matrices. The Cholesky decomposition now depends on the ordering of the variables and hence we compute the IS measure for two system configurations. The first considers the exchange rate and the (preferred) share price at the home market as the first and last variables of the system, respectively. This entails a upper bound for the IS of the exchange rate and a lower bound for the IS of the (preferred) share price at the home market. The second configuration inverts the roles of these two variables and hence gives way to a lower bound for the IS of the exchange rate and a upper bound for the IS of the home price. We find a considerable gap between the lower and upper bounds of the Cholesky-based IS estimates. Averaging the bounds (or across all possible permutations) improves the performance, but not enough to get closer to the true IS values. As we increase the correlation between the markets (i.e., from Ω_2 to Ω_3), the problem becomes even more severe, with the Choleski decomposition rendering very dissimilar information shares according to the ordering of the variables. This confirms that incorrectly imposing a lower-triangular structure for the system is potentially very damaging for a price discovery analysis. In stark contrast, the eigendecomposition renders unique IS estimates that are pretty close to the corresponding true values.

4.5 Conclusion

We conduct a price discovery analysis for dual-class shares that trade at different markets. In particular, we focus on the common and preferred shares

of Petrobras and Vale at the BM&FBovespa and their ADR counterparts at the NYSE. Once we account for the BRL/USD exchange rate, this leads to a system with 5 variables for Vale and 7 variables for Petrobras given that we also observe transactions at the NYSE Arca for the latter. We gauge the contribution of each share class and market by means of Hasbrouck's (1995) information share measure. Unfortunately, the standard framework does not work well for large systems because the Cholesky decomposition it employs imposes *ex-ante* restrictions on which share class and market leads the price discovery process. To circumvent such a constraint, one would have to average the IS measures across all possible permutations of the variables that integrate the system. We thus develop an alternative IS measure that rests on the eigendecomposition of the covariance matrix of the reduced-form errors. In stark contrast to the Cholesky decomposition, the spectral-based approach is order invariant and hence corresponds to an agnostic price discovery analysis that imposes no *a priori* lead-lag pattern in the price dynamics.

Examining both common and preferred shares allows us not only to gather more information about the fundamental value of the company, but also say something about the dual-class premium. The evidence we uncover for Petrobras and Vale are compatible either with the expropriation of preferred shareholders as a class or with the majority shareholder extracting private benefits from their control rights. In both cases, we identify the Brazilian government as the main beneficiary of the dual-class premium. It detains not only Petrobras' control by holding over 55% of the voting shares, but also Vale's indirect control through a consortium of state pension funds. Note that the dual-class premium is a common factor governing

the dynamics of the system only in normal times given that it becomes stationary in periods of financial distress. We also find that the foreign market is more important than the home market for the price discovery in both Petrobras and Vale. As a matter of fact, we notice that the IS estimates we obtain are by a long chalk increasing with the trade intensity of the corresponding price and hence the dominance of the NYSE. This pattern actually becomes more pronounced in the aftermath of the financial crisis, with the BM&FBovespa losing much of its importance for Petrobras and Vale in this period.

As for the exchange rate, we observe that it is the ADR prices that incorporate any shock in the efficient exchange rate. Our results also indicate that the efficient exchange rate is not exogenous to changes in the fundamental values of Petrobras and Vale. We conjecture that this is an artifact due to the omission of commodity indices in the analysis. For instance, one could include international oil prices in the Petrobras' system and the S&P industrial metals spot index in the Vale's analysis. The correlation between changes in commodity prices and the exchange rate variation is normally very high and hence we predict that augmenting the systems would help recover the expected exogeneity of the exchange rate.

Table 4.1
Information share for Petrobras in 2008

We report the IS estimates based on the eigendecomposition of the covariance matrix of the reduced-form errors, and their bootstrap-based standard errors. The first subsample covers 87,781 observations from the first half of 2008, whereas the second subsample has 90,431 observations from the second half of 2008. BRLUSD refers to the exchange rate, PETR3 and PETR4 are the common and preferred shares of Petrobras at the BM&FBovespa, PBR and PBRa are the common and preferred ADRs of Petrobras. The extensions N and P are for NYSE and Arca, respectively.

	January to June			information share					cointegrating vector		
	BRLUSD	PETR4	PETR3	PBR.N	PBRa.N	PBR.P	PBRa.P				
BRLUSD	0.53 (0.027)	0.00 (0.002)	0.04 (0.007)	0.05 (0.008)	0.07 (0.009)	0.16 (0.017)	0.15 (0.014)	-0.02	0.00	0.01	1.00
PETR4	0.00 (0.001)	0.38 (0.025)	0.03 (0.007)	0.12 (0.015)	0.10 (0.012)	0.25 (0.023)	0.11 (0.015)	0.02	0.00	1.00	-0.01
PETR3	0.00 (0.002)	0.23 (0.019)	0.06 (0.010)	0.20 (0.023)	0.06 (0.008)	0.39 (0.026)	0.06 (0.010)	0.00	0.00	-1.00	-1.00
PBR.N	0.02 (0.004)	0.16 (0.017)	0.03 (0.006)	0.20 (0.023)	0.07 (0.010)	0.42 (0.027)	0.09 (0.014)	1.00	0.00	1.01	1.00
PBRa.N	0.04 (0.006)	0.26 (0.021)	0.01 (0.004)	0.13 (0.016)	0.12 (0.014)	0.29 (0.024)	0.15 (0.017)	0.00	1.00	-1.00	0.00
PBR.P	0.02 (0.004)	0.16 (0.017)	0.03 (0.006)	0.20 (0.023)	0.07 (0.010)	0.42 (0.027)	0.09 (0.014)	-1.00	0.00	0.00	-0.04
PBRa.P	0.04 (0.006)	0.26 (0.021)	0.01 (0.004)	0.13 (0.016)	0.12 (0.014)	0.29 (0.024)	0.15 (0.017)	-0.02	-1.00	0.00	0.04
	July to December			information share					cointegrating vector		
	BRLUSD	PETR4	PETR3	PBR.N	PBRa.N	PBR.P	PBRa.P				
BRLUSD	0.38 (0.071)	0.04 (0.027)	0.00 (0.006)	0.09 (0.045)	0.04 (0.029)	0.29 (0.105)	0.14 (0.082)	0.00	0.00	1.00	-0.04
PETR4	0.04 (0.016)	0.14 (0.056)	0.05 (0.017)	0.17 (0.053)	0.07 (0.037)	0.36 (0.105)	0.17 (0.094)	0.00	0.00	-0.01	0.00
PETR3	0.04 (0.015)	0.14 (0.057)	0.05 (0.017)	0.17 (0.053)	0.07 (0.037)	0.36 (0.105)	0.17 (0.093)	0.00	0.00	-1.00	-0.01
PBR.N	0.00 (0.007)	0.05 (0.032)	0.02 (0.013)	0.19 (0.057)	0.08 (0.041)	0.45 (0.114)	0.21 (0.107)	1.00	0.00	0.00	1.00
PBRa.N	0.01 (0.007)	0.05 (0.031)	0.02 (0.013)	0.19 (0.057)	0.08 (0.041)	0.45 (0.114)	0.21 (0.108)	0.00	1.00	0.00	-1.00
PBR.P	0.00 (0.007)	0.05 (0.032)	0.02 (0.013)	0.19 (0.057)	0.08 (0.041)	0.45 (0.114)	0.21 (0.107)	-1.00	0.00	1.00	0.00
PBRa.P	0.01 (0.007)	0.05 (0.031)	0.02 (0.013)	0.19 (0.057)	0.08 (0.041)	0.45 (0.114)	0.21 (0.108)	0.00	-1.00	0.00	-0.04

Table 4.2
Information share for Petrobras in 2009

We report the IS estimates based on the spectral decomposition of the covariance matrix of the reduced-form errors, and their bootstrap-based standard errors. The first subsample covers 87,797 observations from the first half of 2009, whereas the second subsample has 75,010 observations from July to November 2009. BRLUSD refers to the exchange rate, PETR3 and PETR4 are the common and preferred shares of Petrobras at the BM&FBovespa, PBR and PBRa are the common and preferred ADRs of Petrobras. The extensions N and P are for NYSE and Arca, respectively.

	January to June			information share						cointegrating vector					
	BRLUSD	PETR4	PETR3	PBR.N	PBRa.N	PBR.P	PBRa.P								
BRLUSD	0.39 (0.042)	0.00 (0.017)	0.02 (0.004)	0.10 (0.022)	0.10 (0.031)	0.14 (0.044)	0.25 (0.056)	0.00	0.00	1.00	1.00	0.00	0.00	1.00	0.00
PETR4	0.00 (0.002)	0.08 (0.053)	0.05 (0.008)	0.26 (0.037)	0.06 (0.024)	0.41 (0.078)	0.14 (0.042)	0.00	0.00	-0.03	-1.00	1.00	0.00	-1.00	1.00
PETR3	0.00 (0.004)	0.09 (0.056)	0.06 (0.010)	0.26 (0.037)	0.06 (0.022)	0.41 (0.078)	0.12 (0.038)	0.00	0.00	-1.00	-0.01	-1.00	0.00	-1.00	-1.00
PBR.N	0.03 (0.010)	0.06 (0.048)	0.02 (0.005)	0.24 (0.037)	0.08 (0.027)	0.38 (0.075)	0.18 (0.047)	1.00	0.00	0.00	0.00	0.00	1.00	0.00	0.00
PBRa.N	0.06 (0.016)	0.05 (0.044)	0.01 (0.004)	0.23 (0.035)	0.09 (0.029)	0.35 (0.072)	0.21 (0.050)	0.00	1.00	0.00	0.00	0.00	0.00	0.00	0.00
PBR.P	0.03 (0.010)	0.06 (0.048)	0.02 (0.005)	0.24 (0.037)	0.08 (0.027)	0.38 (0.075)	0.18 (0.047)	-1.00	0.00	1.00	0.00	0.00	0.00	0.00	0.97
PBRa.P	0.06 (0.016)	0.05 (0.044)	0.01 (0.004)	0.23 (0.035)	0.09 (0.029)	0.35 (0.072)	0.21 (0.050)	0.00	-1.00	0.02	1.00	0.00	-0.97	1.00	-0.97
	July to November			information share						cointegrating vector					
	BRLUSD	PETR4	PETR3	PBR.N	PBRa.N	PBR.P	PBRa.P								
BRLUSD	0.47 (0.072)	0.00 (0.010)	0.01 (0.011)	0.13 (0.015)	0.09 (0.017)	0.15 (0.019)	0.15 (0.021)	0.01	0.00	0.00	0.00	1.00	0.01	0.00	1.00
PETR4	0.02 (0.014)	0.09 (0.046)	0.01 (0.005)	0.21 (0.014)	0.18 (0.017)	0.24 (0.017)	0.25 (0.018)	-0.01	0.00	1.00	-0.17	0.00	-0.01	0.00	-0.17
PETR3	0.03 (0.016)	0.03 (0.025)	0.02 (0.013)	0.33 (0.017)	0.09 (0.014)	0.37 (0.018)	0.13 (0.015)	0.00	0.00	0.00	-1.00	-1.00	0.00	0.00	-1.00
PBR.N	0.01 (0.008)	0.02 (0.019)	0.01 (0.007)	0.32 (0.017)	0.12 (0.015)	0.36 (0.018)	0.18 (0.016)	1.00	0.00	0.00	1.01	1.00	1.00	0.00	1.00
PBRa.N	0.02 (0.012)	0.05 (0.033)	0.00 (0.005)	0.22 (0.015)	0.19 (0.018)	0.25 (0.017)	0.27 (0.019)	0.00	1.00	-1.00	0.00	0.00	0.00	1.00	0.00
PBR.P	0.01 (0.008)	0.02 (0.019)	0.01 (0.007)	0.32 (0.017)	0.12 (0.015)	0.36 (0.018)	0.18 (0.016)	-1.00	0.00	0.00	0.00	0.03	-1.00	0.00	0.03
PBRa.P	0.02 (0.012)	0.05 (0.033)	0.00 (0.005)	0.22 (0.015)	0.19 (0.018)	0.25 (0.017)	0.27 (0.019)	0.01	-1.00	0.00	0.00	0.06	0.01	-1.00	0.06

Table 4.3
Information share for Vale

We report the IS estimates based on the spectral decomposition of the covariance matrix of the reduced-form errors, and their bootstrap-based standard errors. There are 87,966 observations in the first half of 2008, 90,143 in the second half of 2008, 86,256 in the first half of 2009, and 76,311 observations from July to November 2009. BRLUSD refers to the exchange rate, VALE3 and VALE5 are the common and preferred shares of Vale at the BM&FBovespa, RIO.N and RIOp.N are the common and preferred ADRs of Vale at the NYSE.

January to June 2008		information share				cointegrating vector	
	BRLUSD	VALE5	VALE3	RIO.N	RIOp.N		
BRLUSD	0.42 (0.039)	0.01 (0.006)	0.00 (0.002)	0.21 (0.021)	0.36 (0.029)	1.00	0.00
VALE5	0.00 (0.003)	0.24 (0.028)	0.14 (0.019)	0.22 (0.020)	0.41 (0.029)	0.04	1.00
VALE3	0.00 (0.003)	0.10 (0.018)	0.25 (0.026)	0.42 (0.026)	0.23 (0.023)	-1.05	-0.99
RIO.N	0.02 (0.007)	0.06 (0.014)	0.18 (0.022)	0.45 (0.026)	0.30 (0.026)	1.02	0.99
RIOp.N	0.02 (0.008)	0.14 (0.022)	0.09 (0.017)	0.27 (0.021)	0.48 (0.031)	0.00	-0.99
July to December 2008		information share				cointegrating vector	
	BRLUSD	VALE5	VALE3	RIO.N	RIOp.N		
BRLUSD	0.54 (0.050)	0.04 (0.032)	0.01 (0.006)	0.32 (0.053)	0.10 (0.070)	1.00	-1.02
VALE5	0.00 (0.002)	0.03 (0.032)	0.23 (0.026)	0.71 (0.064)	0.03 (0.034)	0.00	1.00
VALE3	0.00 (0.005)	0.04 (0.036)	0.24 (0.027)	0.69 (0.065)	0.02 (0.030)	-0.99	0.00
RIO.N	0.05 (0.016)	0.01 (0.018)	0.17 (0.022)	0.72 (0.059)	0.05 (0.047)	0.99	0.00
RIOp.N	0.09 (0.022)	0.00 (0.014)	0.14 (0.021)	0.71 (0.058)	0.06 (0.051)	0.00	-1.01
January to June 2009		information share				cointegrating vector	
	BRLUSD	VALE5	VALE3	RIO.N	RIOp.N		
BRLUSD	0.40 (0.043)	0.04 (0.044)	0.01 (0.012)	0.26 (0.051)	0.29 (0.083)	1.00	-1.00
VALE5	0.01 (0.006)	0.00 (0.020)	0.27 (0.035)	0.48 (0.064)	0.23 (0.076)	0.00	1.00
VALE3	0.02 (0.009)	0.01 (0.022)	0.29 (0.036)	0.47 (0.064)	0.22 (0.072)	-1.02	0.00
RIO.N	0.01 (0.004)	0.00 (0.016)	0.22 (0.032)	0.49 (0.063)	0.28 (0.085)	1.01	0.00
RIOp.N	0.02 (0.009)	0.00 (0.017)	0.19 (0.031)	0.49 (0.062)	0.30 (0.089)	0.00	-1.00
July to November 2009		information share				cointegrating vector	
	BRLUSD	VALE5	VALE3	RIO.N	RIOp.N		
BRLUSD	0.49 (0.046)	0.00 (0.006)	0.01 (0.012)	0.19 (0.020)	0.31 (0.045)	1.00	0.00
VALE5	0.00 (0.003)	0.19 (0.041)	0.08 (0.029)	0.28 (0.022)	0.44 (0.052)	0.15	1.00
VALE3	0.00 (0.005)	0.08 (0.026)	0.11 (0.035)	0.49 (0.027)	0.31 (0.042)	-1.36	-1.02
RIO.N	0.02 (0.009)	0.05 (0.021)	0.09 (0.032)	0.48 (0.026)	0.36 (0.045)	1.13	0.99
RIOp.N	0.04 (0.011)	0.11 (0.031)	0.07 (0.027)	0.33 (0.022)	0.46 (0.052)	0.00	-0.97

Table 4.4**Information share for M_1 with a diagonal covariance matrix**

We report the mean estimates of the information share using the spectral and Cholesky decompositions as well as their standard errors within parentheses. All results rest on 1,000 samples of 10,000 observations of model M_1 , fixing the covariance matrix of the errors to identity.

	true			spectral			Cholesky		
	e	p^f	p^h	e	p^f	p^h	e	p^f	p^h
e	1	0	0	1.00 (0.002)	0.00 (0.001)	0.00 (0.002)	1.00 (0.002)	0.00 (0.001)	0.00 (0.002)
p^f	0.5	0	0.5	0.50 (0.032)	0.00 (0.001)	0.50 (0.032)	0.50 (0.032)	0.00 (0.002)	0.50 (0.032)
p^h	0	0	1	0.00 (0.002)	0.00 (0.002)	1.00 (0.002)	0.00 (0.002)	0.00 (0.002)	1.00 (0.002)

Table 4.5
Information share for M_1 with a nondiagonal covariance matrix

We report the mean estimates of the information share using the spectral decomposition as well as the average lower and upper bounds of the Cholesky-based IS estimates relative to the exchange rate, with their standard errors within parentheses. We also inform the mean and standard error of the midpoint between the lower and upper bounds. All results rest on 1,000 samples of 10,000 observations of model M_1 , with the covariance matrix of the errors given by Ω_1 .

	true			spectral			Cholesky (midpoint)		
	e	p^f	p^h	e	p^f	p^h	e	p^f	p^h
e	0.96	0.04	0.00	0.96 (0.013)	0.04 (0.013)	0.00 (0.002)	0.91 (0.013)	0.08 (0.012)	0.01 (0.003)
p^f	0.46	0.10	0.45	0.46 (0.031)	0.10 (0.019)	0.45 (0.031)	0.46 (0.030)	0.10 (0.018)	0.44 (0.029)
p^h	0.00	0.07	0.93	0.00 (0.002)	0.07 (0.016)	0.93 (0.016)	0.01 (0.004)	0.13 (0.014)	0.87 (0.014)

	true			Cholesky (upper)			Cholesky (lower)		
	e	p^f	p^h	e	p^f	p^h	e	p^f	p^h
e	0.96	0.04	0.00	1.00 (0.029)	0.00 (0.029)	0.00 (0.007)	0.82 (0.002)	0.16 (0.001)	0.01 (0.001)
p^f	0.46	0.10	0.45	0.55 (0.030)	0.12 (0.021)	0.34 (0.031)	0.37 (0.031)	0.08 (0.017)	0.55 (0.031)
p^h	0.00	0.07	0.93	0.01 (0.001)	0.25 (0.001)	0.74 (0.002)	0.00 (0.007)	0.00 (0.025)	1.00 (0.025)

Table 4.6
Information share for M_2 with a nondiagonal covariance matrix

We report the mean estimates of the information share using the spectral decomposition as well as the average lower and upper bounds of the Cholesky-based IS estimates relative to the exchange rate, with their standard errors within parentheses. We also inform the mean and standard error of the midpoint between the lower and upper bounds. All results rest on 1,000 samples of 10,000 observations of model M_2 , with the covariance matrix of the errors given by Ω_2 .

true			spectral			Cholesky (midpoint)		
e	p_c^h	p_p^f	p_c^f	p_p^h	e	p_c^h	p_p^f	p_p^h
e	0.94	0.00	0.02	0.04	0.00	0.90	0.00	0.01
	(0.015)	(0.002)	(0.009)	(0.012)	(0.003)	(0.013)	(0.009)	(0.004)
p_c^h	0.00	0.92	0.00	0.04	0.04	0.01	0.91	0.00
	(0.002)	(0.018)	(0.004)	(0.013)	(0.012)	(0.004)	(0.013)	(0.001)
p_p^f	0.46	0.02	0.05	0.03	0.44	0.46	0.04	0.02
	(0.031)	(0.010)	(0.014)	(0.011)	(0.032)	(0.030)	(0.013)	(0.005)
p_c^f	0.45	0.44	0.02	0.07	0.02	0.46	0.44	0.02
	(0.032)	(0.032)	(0.008)	(0.017)	(0.010)	(0.030)	(0.031)	(0.007)
p_p^h	0.00	0.04	0.04	0.00	0.91	0.01	0.08	0.00
	(0.002)	(0.013)	(0.013)	(0.004)	(0.019)	(0.004)	(0.012)	(0.003)

true			Cholesky (upper)			Cholesky (lower)		
e	p_c^h	p_p^f	p_c^f	p_p^h	e	p_c^h	p_p^f	p_p^h
e	0.94	0.00	0.02	0.04	0.00	0.81	0.01	0.08
	(0.003)	(0.001)	(0.001)	(0.001)	(0.002)	(0.025)	(0.005)	(0.018)
p_c^h	0.00	0.92	0.00	0.04	0.00	0.00	0.84	0.00
	(0.007)	(0.007)	(0.001)	(0.002)	(0.001)	(0.002)	(0.025)	(0.001)
p_p^f	0.46	0.02	0.05	0.03	0.33	0.37	0.00	0.04
	(0.032)	(0.017)	(0.014)	(0.004)	(0.030)	(0.032)	(0.003)	(0.012)
p_c^f	0.45	0.44	0.02	0.07	0.00	0.37	0.44	0.04
	(0.032)	(0.032)	(0.001)	(0.002)	(0.002)	(0.031)	(0.033)	(0.012)
p_p^h	0.00	0.04	0.04	0.00	0.72	0.00	0.00	0.00
	(0.007)	(0.024)	(0.021)	(0.005)	(0.031)	(0.002)	(0.001)	(0.002)

Table 4.7
Information share for M_3 with a nondiagonal covariance matrix

We report the mean estimates of the information share using the spectral decomposition as well as the average lower and upper bounds of the Cholesky-based IS estimates relative to the exchange rate, with their standard errors within parentheses. We also inform the mean and standard error of the midpoint between the lower and upper bounds. All results rest on 1,000 samples of 10,000 observations of model M_3 , with the covariance matrix of the errors given by Ω_3 .

true			spectral			Cholesky (midpoint)		
e	p_c^h	p_p^f	p_c^f	p_p^h	e	p_c^h	p_p^f	p_p^h
e	0.87	0.04	0.04	0.05	0.04	0.80	0.01	0.02
	(0.021)	(0.014)	(0.014)	(0.014)	(0.013)	(0.016)	(0.004)	(0.010)
p_c^h	0.04	0.19	0.09	0.09	0.09	0.13	0.28	0.00
	(0.013)	(0.026)	(0.018)	(0.018)	(0.033)	(0.015)	(0.017)	(0.001)
p_p^f	0.43	0.14	0.08	0.09	0.03	0.47	0.10	0.02
	(0.032)	(0.023)	(0.018)	(0.018)	(0.031)	(0.026)	(0.014)	(0.006)
p_c^f	0.43	0.14	0.08	0.09	0.03	0.47	0.10	0.02
	(0.032)	(0.023)	(0.018)	(0.018)	(0.031)	(0.026)	(0.014)	(0.006)
p_p^h	0.04	0.19	0.09	0.09	0.09	0.13	0.28	0.00
	(0.013)	(0.026)	(0.018)	(0.018)	(0.033)	(0.015)	(0.017)	(0.001)
true			Cholesky (upper)			Cholesky (lower)		
e	p_c^h	p_p^f	p_c^f	p_p^h	e	p_c^h	p_p^f	p_p^h
e	0.87	0.04	0.04	0.00	1.00	0.60	0.01	0.04
	(0.003)	(0.001)	(0.001)	(0.002)	(0.031)	(0.031)	(0.008)	(0.013)
p_c^h	0.04	0.19	0.09	0.01	0.25	0.00	0.00	0.00
	(0.030)	(0.033)	(0.006)	(0.002)	(0.001)	(0.001)	(0.002)	(0.001)
p_p^f	0.43	0.14	0.08	0.00	0.74	0.20	0.01	0.01
	(0.029)	(0.026)	(0.004)	(0.002)	(0.025)	(0.025)	(0.005)	(0.008)
p_c^f	0.43	0.14	0.08	0.00	0.74	0.20	0.01	0.01
	(0.029)	(0.026)	(0.004)	(0.002)	(0.025)	(0.025)	(0.005)	(0.008)
p_p^h	0.04	0.19	0.09	0.01	0.25	0.00	0.00	0.00
	(0.030)	(0.033)	(0.006)	(0.002)	(0.001)	(0.001)	(0.002)	(0.001)

Chapter 5

A Dynamic Measure for Price Discovery

5.1 Introduction

Financial markets can be very informative of the economic situation of a business, an industry, a country. The interaction among players for allocation of capital ownership defines financial prices. Assets price changes reflect how new information is incorporated into perceptions, beliefs and assumptions. Prices at different markets may react differently to news, as a consequence of market's structure, design, liquidity, efficiency, behaviour. We study how assets driven by the same fundamentals and traded at various markets react to news. We show that price's fundamentals may have cross linkages, meaning that an innovation on exchange rate may have a permanent effect on the asset latent price, for instance. This is a major contribution to the price discovery literature, since no study has considered cross linkages among common factors before.

We compute a dynamic price discovery measure on common and preferred shares ¹ traded at both domestic and foreign markets. We set up a novel market microstructure model allowing for cross linkages among common factors. In this model, we isolate instantaneous and long-run effects on prices given structural innovations associated to common factors. Our flexible econometric specification, differently from previous dynamics measures in price discovery, allows us to look at cross linkages in the same way the theoretical model does. We show how to identify the structural innovations with minimal restrictions. Our theoretical model and econometric methodology yield three developments: First, we obtain dynamic measures of price discovery that quantify how prices traded at different venues respond to shocks on common factors over time. Second, we are able to test whether a devaluation of a currency leads to a devaluation of the asset exceeding the expected arbitrage adjustment. Third, we test whether changes on the latent voting premium have effect not only on common shares, but also on preferred ones. These results in a novel conclusion on the dynamic price discovery analyses: an innovation of one unit on common factors may have a larger impact on observed prices, given cross linkages among common factors.

Static price discovery has been a subject widely studied. The two most prominent static measures of price discovery are the information share (IS) and component share (CS) frameworks of Hasbrouck (1995) and Gonzalo and Granger (1995) respectively. These two methodologies and their numerous variations were broadly applied to different markets, assets and financial instruments. These studies have mainly focused on identifying

¹Preferred shares have preference on receiving dividends and do not hold voting rights.

either the market or the financial instrument that is the fastest on impounding new information. Harris, McNish, and Wood (2002) measure price discovery for stocks in the Dow Jones Industrial Average (DJIA) index traded at different exchanges and find that NYSE is “information dominant”. Chu, Hsieh, and Tse (1999) and Booth, So, and Tseh (1999) find that future markets are the leaders in impounding new information. Hasbrouck (2003) shows that E-mini contracts and exchange-traded index mutual funds (ETFs) are the fastest ones in impounding new information. More recently, Grammig, Melvin, and Schlag (2005) introduced the exchange rate on the price discovery analysis using the IS framework. They focus on three German firms cross listed in the US market and their results strongly suggest that the home market is the one playing the main role on the price discovery process. By adopting a variant of the framework suggested by Grammig, Melvin, and Schlag (2005) and introducing a unique IS measure, Chapter 4 finds that the relative importance of the home and foreign markets may change over time, specially given financial crises events.

The main drawback associated with the two static methodologies is that both are based on market innovations in their reduced form. Lehmann (2002) points out that price discovery is a dynamic process and that both IS and CS measures are not able to capture the price dynamics. Fundamentally, shocks on observed prices may be correlated. Specifically to measure price discovery, one would be interested in looking at instantaneous and total effects of uncorrelated innovations. Therefore, in order to appropriately address the price discovery dynamics it is paramount to adopt a structural methodology such that changes in the structural innova-

tions can be correctly assigned to markets. Yan and Zivot (2007) introduce the first structural measure of price discovery. Their framework considers only one common factor, and they find that price discovery of the yen/euro exchange rate happens mainly through the US dollar market.

In our market microstructure model, we include latent and observed prices. Observed prices are allowed to share not only one common factor, but either two or three, namely, the efficient price, the efficient exchange rate, and the efficient voting premium. The innovations attached to each one of the three common factors are allowed to be contemporaneously correlated. This provides a considerable advantage to our theoretical approach, since it encompasses results such as cross linkages among the common factors in the short- and long-run, as well as on the observed prices. By using the structural framework, we are able to retrieve the structural counterpart associated to each common factors innovation. These structural innovations have a diagonal covariance matrix, allowing to explicitly obtain the effects of shocks on each of the observed prices across every period of time. We present in detail the short- and long-run solutions for a five-variable model.

With regard to the estimation process, we merge the two-step procedure proposed by Gonzalo and Ng (2001) and the methodology suggested by Warne (1993) in order to recover the permanent and transitory shocks in their reduced form. As a result, we do not have to face normalization issues, regarding the cointegrating vectors. To obtain the structural counterparts, we decompose the covariance matrix of the reduced form innovations allowing for the structural innovations variance to be different than one. Moreover, we chose a decomposition flexible enough to deliver

any shape of relation between reduced and structural innovations, differently than the decompositions proposed in the price discovery literature. Combining both methodological contributions, we find that our measure is order invariant. Moreover, we present a Monte Carlo exercise showing that our measure outperforms, in terms of relative mean squared error (RMSE), the measures proposed in Gonzalo and Ng (2001) and Yan and Zivot (2007). Yan and Zivot (2007) apply the Gonzalo and Ng (2001) approach to the price discovery study for a two-variable model, also allowing for the structural variance matrix to be different from the identity matrix. Kim (2010a) expands the work of Yan and Zivot (2007) to a three-variable model sharing two common factors (only one cointegrating vector). Our work differs from theirs in three ways: first, we introduce a theoretical model where we allow for contemporaneous correlation among common factors, and hence, short- and long-run solutions are identified as functions of the parameters that relate correlated innovations to structural innovations. Second, we decompose the covariance matrix of the reduced innovations in such a way that common factors may impact all observed prices in the long-run. This is an important advantage of our approach when compared to the standard triangular decomposition adopted in Yan and Zivot (2007) and Kim (2010a). Under their approach the sub-matrix containing the long-run effects of changes in the structural permanent innovations is restricted to have ones in its diagonal and zeros on the upper block. Third, as discussed in the Monte Carlo section, our methodology contribution overcomes the estimation problem that arises in their work when models have more than one cointegrating vector (the number of cointegrating vectors increases as one adds more variables sharing the same common factor to our model,

since there are additional combinations of variables delivering a stationary process). Furthermore, we show that our theoretical results and methodology holds for a seven-variable model with three common factors.

We study the instantaneous and long-run price effects for cross listed Brazilian firms traded at Brazil and US. We work with a tick by tick data set from BM&FBovespa (the Brazilian stock exchange), NYSE and ARCA incorporating common and preferred shares. Brazilian firms are traded at the US markets as American Depositary Receipts (ADR). The data ranges from December 2007 to November 2009. There are important characteristics that make this data set relevant for this study. The first one is regarding core trading hours. There is a large trade intersection period in the entire year between Brazil and US, varying from six hours and thirty minutes to five hours and thirty minutes. This allows us to have much more information regarding the price discovery process, when compared to studies among companies traded at the European and the US markets, where the intersection period is merely two to three hours. Second, we work with companies that possess distinctive characteristics such as: core business, ownership structure, global insertion, and strategic and political relevance. Brazilian firms are Ambev (beverage, private owned, global market), BR Telecom (telecommunication, private owned, domestic market), Bradesco (finance, private owned, domestic market), Gerdau (steel, private owned, global company), Vale (mining, private owned with governmental influence, global company) and Petrobras (oil, state owned, global company). Apart from BR Telecom, they all are part of Ibovespa, the main index of the Brazilian stock exchange. Vale and Petrobras are the largest Brazilian companies and this is reflected in their weights on the Ibovespa index.

Liquidity plays an important role on defining whether both common and preferred shares traded in all exchanges can be considered in the baseline model. We require stocks to be highly liquid as a condition to be part of the analysis. This implies that the number of variables varies from company to company. We report results considering four-, five- and seven-variable models. In all different specifications, we are able to understand which markets are the main drivers for the price determination of these Brazilian cross listed firms. Moreover, by including the exchange rate in all models, we test whether shocks on the innovations associated with the efficient price and efficient exchange rate have a long-run impact on the exchange rate and firm's value, respectively. The five- and seven-variable models include the common shares traded at both domestic and foreign markets. This adds the voting premium as a further common factor to the price dynamics analysis. Under this specification, we test whether shocks on the voting premium lead to a permanent effect on the preferred shares.

The empirical results corroborate the solutions of our market microstructure model, implying that innovations associated with the latent processes are contemporaneously correlated, leading to cross linkages among common factors. This shows that measuring price discovery independently of exchange rate or other common factors may lead to misleading results. We document that, in the long-run, a depreciation of the Brazilian currency leads to a depreciation of the value of the firm that exceeds the expected arbitrage adjustment. In addition, a positive shock on the voting premium yields a positive impact on the value of the firm. In general, ARCA is faster than NYSE in the short run, but they are equally important in the long-run. These results are consistent across all the six companies, as well

as at different sampling frequencies. Our price discovery analysis also reveals that one trading day suffices to impound new information on all share prices, regardless of the venue they trade at

The remaining of the chapter proceeds as follows: Section 5.2 introduces the theoretical model. Section 5.3 describes the estimation procedure and show our identification strategy. Section 5.4 documents the empirical results for Brazilian firms. Section 5.6 presents additional results regarding our identification strategy and Section 5.7 has the Monte Carlo study addressing the performance of our estimation methodology. Section 5.8 offers some concluding remarks.

5.2 A simple model for price discovery

We present a simple price discovery model, in order to guide the understanding of the empirical results. We consider a firm traded at four markets and the exchange rate. There is a common and preferred share in both the home market and the foreign market. This model setup can be easily extended to the case with six markets plus the exchange rate and also reduced to the case of three markets plus exchange rate case (as is the instance for some companies analyzed in the empirical section).

The main target of this model is to have price variations in the short run (instantaneous effects) and in the long-run as a function of permanent and transitory uncorrelated innovations, as presented in Gonzalo and Granger (1995) and Gonzalo and Ng (2001) and explained in detail in the next section. By implementing this breakdown, we are able to isolate permanent innovations coming from different sources. We have instantaneous effects

when $L = 0$ (L being the lag operator). This is the effect at the same period in time of the innovation. This comes from a Vector Moving Average (VMA) Model, where the matrix giving instantaneous effects into prices is different than the identity matrix (structural form). As long-run, we refer to $L = 1$, which gives the sum of impacts from innovations across time in a VMA model. This measure gives the total effect on prices of an innovation.

We write the efficient price and efficient exchange rate as random walk processes being affected by permanent innovations. We insert a third efficient factor that is common to the observed prices, the efficient voting premium. We also model it as a random walk process, since from the empirical results, we find systems sharing three common factors, leading to the conclusion that the voting premium is a non-stationary random walk process. Additionally, we could not find any plausible reason for the voting premium not be a random walk. It is a financial asset at last, and as so, it follows the unpredictability characteristic on its returns. These three random walks are non observable prices. Each permanent innovation related to a particular common factor may affect other non observable prices. This effect is given by λ , ρ , π and κ . In other words, innovations on common factors are correlated. We aim to isolate them and quantify their impacts in each of the observed prices. The intuition on the allowance for this correlation comes from empirical analysis of the data. For instance, the correlation between the exchange rate (Brazilian currency over US dollars) and the main index of the Brazilian stock exchange (Ibovespa) is equal to -0.60 during December 2007 and November 2009. These are observed prices, but brings the questions if the same is true for the latent process of these prices. If there were no cross linkages among the common factors,

one would expect to find these parameters equal to zero in the empirical results. If markets are instantaneously efficient, the difference between each of these prices and their respective observed counterparts are the transitory effects at each point in time. For instance, the observed price is equal to the efficient price plus transitory effects, such as bid ask bounces, price discreteness, inventory effects, etc. In the same way that the observed voting premium (the difference between the common and the preferred share, considering efficiency of these observable prices on incorporating news on the efficient price) is equal to the efficient voting premium plus transitory innovations, such as liquidity effects.

The permanent innovations, η_t , are defined with $\mathbb{E}(\eta_t) = 0$ in their structural form, implying that $\text{Var}(\eta_t)$ is a diagonal matrix. We define the logarithm function of the efficient price of the asset (m_t), of the efficient exchange rate (\dot{e}_t) and of the voting premium (v_t), such that:

$$\dot{e}_t = \dot{e}_{t-1} + \eta_t^e + \lambda \eta_t^m \quad (5.1)$$

$$m_t = m_{t-1} + \eta_t^m + \rho \eta_t^e + \pi \eta_t^v \quad (5.2)$$

$$v_t = v_{t-1} + \eta_t^v + \kappa \eta_t^m \quad (5.3)$$

where η_t^e , η_t^m and η_t^v are the permanent innovations associated to exchange rate, efficient price and voting premium, respectively; $\eta_t = (\eta_t^e, \eta_t^m, \eta_t^v)'$; and \dot{e}_t is defined in terms of home currency. Note that, from the structure imposed in (5.1), (5.2) and (5.3), the efficient price, exchange rate and voting premium are random walk processes, implying that their first difference are

$I(0)$ processes with covariance matrix given by:

$$\begin{aligned} \text{Var}([\Delta \dot{e}_t, \Delta m_t, \Delta v_t]') = \\ = \begin{pmatrix} \varsigma_e^2 + \lambda^2 \varsigma_m^2 & \rho \varsigma_e^2 + \lambda \varsigma_m^2 & \lambda \kappa \varsigma_v^2 \\ \rho \varsigma_e^2 + \lambda \varsigma_m^2 & \varsigma_m^2 + \rho^2 \varsigma_e^2 + \pi^2 \varsigma_e^2 & \kappa \varsigma_v^2 + \pi^2 \varsigma_v^2 \\ \lambda \kappa \varsigma_v^2 & \kappa \varsigma_m^2 + \pi^2 \varsigma_v^2 & \varsigma_v^2 + \kappa^2 \varsigma_m^2 \end{pmatrix} \end{aligned} \quad (5.4)$$

where $\text{Var}(\eta_t^e) = \varsigma_e^2$, $\text{Var}(\eta_t^m) = \varsigma_m^2$ and $\text{Var}(\eta_t^v) = \varsigma_v^2$.

Denote Y_t the vector containing the logarithm function of the exchange rate and observed prices on different venues. We want a high frequency trading model that reflects price adjustments in a partial way, such that innovations are not completely incorporated by all market in each t , i.e. in each microsecond. If markets were efficient, we could have the parameters giving this partial adjustment set to unit, making observed prices equal to efficient price plus some transitory effects, such as bid-ask bounce, price discreteness, liquidity effects, etc. The model below is a modified and extended version of other models used in the literature (see Amihud and Mendelson (1987), Hasbrouck and Ho (1987) and Yan and Zivot (2010)).

We present the observed exchange rate as a function of the efficient exchange rate, past observed exchange rate and transitory effects (denoted by the 2×1 vector η_t^T and the 1×2 vector of parameters b_1). We impose two transitory innovations because we need the number of innovations to be equal to the number of markets. This goes in line with Gonzalo and Ng (2001) and Yan and Zivot (2007), since there is the need to invert a decomposition of the covariance matrix, as the next section explains in detail. Hence, if one has n number of markets, it is necessary to have $n - p$

number of transitory effects, and p number of permanent effects. In this model, we have three permanent innovations (η_t^e , η_t^m and η_t^v), two transitory ones and five variables, allowing invertibility of some specific matrices in our identification strategy. Adjustments to permanent innovations are allowed to happen in a partial way, by inserting γ_i . Hence, all observed prices are allowed to adjust to the three latent prices. We distinguish prices traded at different currencies, such that $y_{4,t}^*$ and $y_{5,t}^*$ entail prices in foreign currency (currency that they are actually traded), whereas $y_{4,t}$ and $y_{5,t}$ are expressed in home currency.

$$y_{4,t}^* = y_{4,t} - e_t$$

$$y_{5,t}^* = y_{5,t} - e_t$$

We write the price process for each $y_{i,t}$ as below.

$$e_t = e_{t-1} + \gamma_1 (m_t - m_{t-1}) + \dot{\gamma}_1 (\dot{e}_t - e_{t-1}) + \tilde{\gamma}_1 (v_t - v_{t-1}) + \mathbf{b}_1 \boldsymbol{\eta}_t^T \quad (5.5)$$

$$y_{2,t} = y_{2,t-1} + \gamma_2 (m_t - y_{2,t-1}) + \dot{\gamma}_2 (\dot{e}_t - e_{t-1}) + \tilde{\gamma}_2 (v_t - v_{t-1}) + \mathbf{b}_2 \boldsymbol{\eta}_t^T \quad (5.6)$$

$$y_{3,t} = y_{3,t-1} + \gamma_3 (m_t - y_{3,t-1} + v_{t-1}) + \dot{\gamma}_3 (\dot{e}_t - e_{t-1}) + \tilde{\gamma}_3 (v_t - v_{t-1}) + \mathbf{b}_3 \boldsymbol{\eta}_t^T \quad (5.7)$$

$$y_{4,t}^* = y_{4,t-1}^* + \gamma_4 (m_t - y_{4,t-1}) + \dot{\gamma}_4 (\dot{e}_t - e_{t-1}) + \tilde{\gamma}_4 (v_t - v_{t-1}) + \mathbf{b}_4 \boldsymbol{\eta}_t^T \quad (5.8)$$

$$y_{5,t}^* = y_{5,t-1}^* + \gamma_5 (m_t - y_{5,t-1} + v_{t-1}) + \dot{\gamma}_5 (\dot{e}_t - e_{t-1}) + \tilde{\gamma}_5 (v_t - v_{t-1}) + \mathbf{b}_5 \boldsymbol{\eta}_t^T \quad (5.9)$$

Where \mathbf{b}_1 , \mathbf{b}_2 , \mathbf{b}_3 , \mathbf{b}_4 and \mathbf{b}_5 are 1×2 vectors and $\boldsymbol{\eta}_t^T$ is a 2×1 vector. We show the steps to obtain $\Delta y_{i,t}$ only for the preferred share traded at the domestic market ($\Delta y_{2,t}$). The remaining equations ((5.5), (5.7), (5.8) and

(5.9) are obtained in a similar manner.

$$\begin{aligned}
y_{2,t} - y_{2,t-1} &= y_{2,t-1} - y_{2,t-2} + \gamma_2 (m_t - m_{t-1} - y_{2,t-1} + y_{2,t-2}) + \mathbf{b}_2 (\boldsymbol{\eta}_t^T - \boldsymbol{\eta}_{t-1}^T) \\
(1 - L + L\gamma_2) \Delta y_{2,t} &= \gamma_2 (\eta_t^m + \rho\eta_t^e + \pi\eta_t^v) + \mathbf{b}_2 (\boldsymbol{\eta}_t^T - L\boldsymbol{\eta}_t^T) \\
\Delta y_{2,t} &= (1 - L + L\gamma_2)^{-1} [\gamma_2 (\eta_t^m + \rho\eta_t^e + \pi\eta_t^v) + \mathbf{b}_2 (\boldsymbol{\eta}_t^T - L\boldsymbol{\eta}_t^T)]
\end{aligned}$$

By setting the lag operator equal to zero, we have the instantaneous effects of permanent and transitory innovations for each price series.

$$\begin{pmatrix} \Delta e_t \\ \Delta y_{2,t} \\ \Delta y_{3,t} \\ \Delta y_{4,t}^* \\ \Delta y_{5,t}^* \end{pmatrix} = \begin{pmatrix} \dot{\gamma}_1 + \gamma_1 \rho & \dot{\gamma}_1 \lambda + \gamma_1 + \tilde{\gamma}_1 \kappa & \tilde{\gamma}_1 + \gamma_1 \pi & \mathbf{b}_1 \\ \dot{\gamma}_2 + \gamma_2 \rho & \dot{\gamma}_2 \lambda + \gamma_2 + \tilde{\gamma}_2 \kappa & \tilde{\gamma}_2 + \gamma_2 \pi & \mathbf{b}_2 \\ \dot{\gamma}_3 + \gamma_3 \rho & \dot{\gamma}_3 \lambda + \gamma_3 + \tilde{\gamma}_3 \kappa & \tilde{\gamma}_3 + \gamma_3 \pi & \mathbf{b}_3 \\ \dot{\gamma}_4 + \gamma_4 \rho & \dot{\gamma}_4 \lambda + \gamma_4 + \tilde{\gamma}_4 \kappa & \tilde{\gamma}_4 + \gamma_4 \pi & \mathbf{b}_4 \\ \dot{\gamma}_5 + \gamma_5 \rho & \dot{\gamma}_5 \lambda + \gamma_5 + \tilde{\gamma}_5 \kappa & \tilde{\gamma}_5 + \gamma_5 \pi & \mathbf{b}_5 \end{pmatrix} \begin{pmatrix} \eta_t^e \\ \eta_t^m \\ \eta_t^v \\ \boldsymbol{\eta}_t^T \end{pmatrix} \quad (5.10)$$

Where \mathbf{b}_1 , \mathbf{b}_2 , \mathbf{b}_3 , \mathbf{b}_4 and \mathbf{b}_5 are 1×2 vectors and $\boldsymbol{\eta}_t^T$ is a 2×1 vector. By making the lag operator equal to unit, we get the long-run effect on prices, as below:

$$\begin{pmatrix} \Delta e_t \\ \Delta y_{2,t} \\ \Delta y_{3,t} \\ \Delta y_{4,t}^* \\ \Delta y_{5,t}^* \end{pmatrix} = \begin{pmatrix} 1 & \lambda & 0 & \mathbf{0} \\ \rho & 1 & \pi & \mathbf{0} \\ \rho & \kappa + 1 & \pi + 1 & \mathbf{0} \\ \rho - 1 & 1 - \lambda & \pi & \mathbf{0} \\ \rho - 1 & 1 + \kappa - \lambda & \pi + 1 & \mathbf{0} \end{pmatrix} \begin{pmatrix} \eta_t^e \\ \eta_t^m \\ \eta_t^v \\ \boldsymbol{\eta}_t^T \end{pmatrix}, \quad (5.11)$$

where $\mathbf{0}$ are 1×2 vectors and $\boldsymbol{\eta}_t^T$ is a 2×1 vector.

To evaluate which market is more important in the price discovery process, we need to look at the combination of parameters from each equation

in the short-run ($L = 0$) and long-run ($L = 1$). When we refer to long-run this is about a day or even less in the high frequency context, differently than in macroeconomic scenarios of years or decades. The combination of parameters is exactly what we estimate and quantify as price discovery measures in the simulations subsection (for a simpler model) and in the empirical results. Let us call the matrix with the combination of parameters that gives the instantaneous effects as D_0 , where on the rows there is each market's effect given an innovation on the permanent and transitory shocks. Similarly, we define $D(1)$ as the effects when $L = 1$. Therefore, we can gauge price discovery by two measures: importance compared to other markets, considering each permanent innovation and fastness, as below:

$$\text{Importance} = \text{Max } d_{0,ii} \text{ and } \text{Max } d(1)_{ii} \quad (5.12)$$

$$\text{Fastness} = d_{0,ii}/d(1)_{ii} \quad (5.13)$$

where $d_{0,ii}$ and $d(1)_{ii}$ are elements of D_0 and $D(1)$ matrices, respectively.

The matrices D_0 and $D(1)$ have dimension $K \times K$, and they give the instantaneous and long-run impacts on market prices from shocks on the permanent and transitory innovations. We are only interested in the parameters accompanying the permanent innovations, because these are the relevant parameters for price discovery. Therefore, we do not look at the part of D_0 or $D(1)$ related to the transitory shocks.

We look at the overall importance when we compare the parameters of D_0 and $D(1)$ for all the markets. The fastness measure may be understood as the proportion of permanent shocks incorporated instantaneously compared to what is incorporated in the long-run. It is important to point

out that both measures depend on the positivity of the elements in D_0 and $D(1)$.

Apart from the price discovery analyses, a question now lies on what to expect from the parameters defined in (5.1), (5.2) and (5.3). We expect to find $\pi > 0$, meaning that a positive innovation on the efficient voting premium delivers a positive impact on the efficient price. If the voting rights of a given firm turn to be more valuable, this would deliver an increase in the value of the asset itself, i.e., the efficient price. An increase in the voting premium is the same as an increase in the price for a vote. Hence, if a vote turns to be more expensive, it is possible that better decisions are taken by the ones owning these rights, once they have paid more for them. Rational individuals should do a better use of something they start paying more for. Therefore, the efficient price is affected positively by the expectations and realizations of better votes.

We conjecture to observe $\rho < 0$, where a positive innovation on the exchange rate (depreciation of the home currency) leads to a negative effect on the efficient price. Ignoring firm specificities regarding imports and exports, a depreciation of the currency where the business is situated and has its main operations would result in a decrease in its value. Exchange rate may affect a firm business in many different fronts: transaction (imports and exports), competitors, suppliers, suppliers competitors, access to international capitals, and so on. This last one, particularly, may impact considerably the cost of firms that aim to finance investments with external resources. This may affect Brazilian companies. As they are located at an emerging country, they do search for external capital resources. A useful literature review on effects of exchange rate on firm value can be found at

Muller and Verschoor (2006).

We would expect λ to be in general equal to zero, since innovations on the efficient price should not lead to effects on the exchange rate. However, depending on how related the firm activities can be to the exchange rate, or how the overall movements of the stock exchange in the home country can be correlated to exchange rate, we might see λ different than zero. We would expect anything different than zero in λ to be related to correlation between observed exchange rate and observed prices.

The observed voting premium (observed difference between common and preferred shares) can be defined as the efficient voting premium plus some transitory effects. These effects can be the result of liquidity issues, such as that either the common or the preferred share has a more liquid market than the other, allowing investors to price this difference. The efficient voting premium is a function of private benefits an investor could get from holding voting rights as well as function of a possible premium over the preferred share, in the instance of a merger or an acquisition (see Zingales (1994) and Zingales (1995) for explanations on private benefits and merger premium, and for reasons on why voting premium vary across countries). Given that, if an increase in the firm's value generates an increase in private benefits or a potential acquisition premium, κ should be positive, leading to a positive impact on the efficient price. This yields a positive effect on the voting premium. However, if we consider that appropriation of private benefits is more related to the culture of the company, and how strong or weak the country institutions are, they should not be affected by the efficient price, unless this innovation in the efficient price is coming from a change in appropriation of private benefits per se.

5.3 Getting structural parameters from reduced-form VECM

Our data set is composed by a single security being traded at different markets, namely Brazil and USA. As they share at least one common factor among them, they are cointegrated. Hence, we use a vector error correction model (VECM) to estimate the price discovery parameters. As we aim to have a structural measure, we want to recover the structural innovations from the VECM residuals. We use Gonzalo and Granger (1995) and Gonzalo and Ng (2001) methodology to retrieve reduced form permanent and transitory innovations from market residuals in their reduced form. We merge their methodology with the work of Warne (1993), making our price discovery measures more accurate. In order to transform reduced form permanent innovations into structural innovations, we modify the standard procedure on Gonzalo and Ng (2001), allowing the variance of the structural innovations to be different than the identity matrix. We show that this methodology works well for models with one, two or three common factors.

Hence, the first step is to estimate a reduced-form VEC model as,

$$\Delta y_t = \xi_1 \Delta y_{t-1} + \xi_2 \Delta y_{t-2} + \dots + \xi_p \Delta y_{t-p} + \zeta + \xi_0 y_{t-1} + \epsilon_t, \quad (5.14)$$

where y_t is a vector of price series in different markets, $\xi_0 = \alpha\beta'$ and ϵ_t is a zero mean white noise process with a non-diagonal covariance matrix Ω . We impose restrictions on the constant term for the absence of deterministic time trends.

Gonzalo and Granger (1995) propose a way to estimate the common fac-

tors from a reduced form model (VECM) and Gonzalo and Ng (2001) show a two-step procedure on how to obtain permanent and transitory structural innovations from the reduced-form errors. To this purpose, we first estimate (5.14) using a the full-information maximum likelihood (FIML) approach proposed by Johansen (1988) and Johansen (1991) and discussed in Hamilton (1994) in order to avoid any possible misspecification in the model derived from setting normalization conditions on the cointegrating vector.

Once we estimate the VEC parameters, we are in position where we can back out the vector moving average (VMA) coefficients through dynamic simulation (see Hamilton (1994)). Note that the VMA equation in (5.15) is driven by the reduced form errors.

$$\Delta y_t = \epsilon_t + \psi_1 \epsilon_{t-1} + \psi_2 \epsilon_{t-2} + \dots = \Psi(L) \epsilon_t, \quad (5.15)$$

In the above equation, change in prices are given by the market reduced form contemporaneous innovations ϵ_t and lagged values. The problem lies on the likely contemporaneous correlation of ϵ_t . Innovations in the home market might be contemporaneously correlated with innovations in the foreign market, making the price discovery analysis problematic. Our target is to have a VMA expression with relation to structural innovations, where the contemporaneous correlation among these innovations are null.

If we were to estimate VMA parameters with relation to market structural innovations, we would have imposed a series of restrictions in order to identify these parameters. These restrictions would require prior knowledge from the markets, which turns to be difficult to be justified. It is less harmful, therefore, to consider assumptions on permanent and transitory

structural innovations since they are easier to implement, not requiring any prior judgement about the markets.

Hence, we want to obtain a VMA expression as in (5.15), that now is driven by the structural permanent and transitory shocks.

$$\Delta y_t = d_0 \eta_t + d_1 \eta_{t-1} + d_2 \eta_{t-2} + \dots = D(L) \eta_t. \quad (5.16)$$

The covariance matrix of $\eta_t = (\eta_t^P, \eta_t^T)'$ is diagonal, where η_t^P entails the permanent effect in y_t and η_t^T consists only of transitory effects. Gonzalo and Granger (1995) define $\lim_{h \rightarrow \infty} \partial E_t(y_{t+h} / \partial \eta_t^{P'}) \neq 0$ and $\lim_{h \rightarrow \infty} \partial E_t(y_{t+h} / \partial \eta_t^{T'}) = 0$, where E_t denotes the conditional expectation with relation to past information up to time t . Therefore the long-run forecast of change in prices can only come permanent innovations. Gonzalo and Ng (2001) rotate ϵ_t and split it into permanent and transitory innovation using matrix G (their first step on finding structural innovations), as below.

$$G = [\alpha'_\perp, \beta']', \quad (5.17)$$

where α_\perp is a $(k - r) \times k$ matrix and β is a $r \times k$ matrix.

To rotate ϵ_t and find the reduced form permanent and transitory innovations, we need to multiply matrix G by ϵ_t as in (5.18),

$$\varepsilon_t = G \epsilon_t, \quad (5.18)$$

where $\varepsilon_t = (\varepsilon_t^P, \varepsilon_t^T)'$. The permanent shock in the reduced form is given by $\alpha'_\perp \epsilon_t$, whereas the transitory shock is $\beta' \epsilon_t$. The variance (Ξ) of the un-

orthogonalized shocks ε_t is still non diagonal, which opens the way for the second step: decompose Ξ in such a way that we find a relation between Ξ and the variance of the structural innovations ($\text{Var}(\eta_t)$). Once we have this, we can extend it to find the relation between ϵ_t and η_t , as well as between the parameters of the VMA in its reduced form and of the structural VMA. Gonzalo and Ng (2001) decompose Ξ using the well known Cholesky decomposition, as below:

$$\Xi = G\Omega G' = CC' = CIC', \quad (5.19)$$

where C is the Cholesky decomposition of the variance matrix of ε_t .

Setting $\text{Var}(\eta_t)$ equal to an identity matrix gives the easy relation between the variance matrices, with $\Xi = CIC' = C\text{Var}(\eta_t)C'$. Now, we are in the position to find the relation between ϵ_t and η_t , given by

$$\eta_t = C^{-1}\varepsilon_t = C^{-1}G\epsilon_t = D_0^{-1}\epsilon_t, \quad (5.20)$$

where $D_0 = G^{-1}C$.

Change in price series are given by the VMA parameters as in (5.15), as well as (5.16). The way to recover the latter one using the estimated parameters from the former is given below.

$$\Delta y_t = \Psi(L)G^{-1}CC^{-1}G\epsilon_t = D(L)\eta_t, \quad (5.21)$$

which delivers $D(L) = \Psi(L)G^{-1}C = \Psi(L)D_0$. Yan and Zivot (2007) use this approach in the price discovery analysis. They implement a modification that allows for the variance matrix of η_t to be different than the

identity matrix, but obviously still diagonal. In particular, they implement a slightly different decomposition than the Cholesky one, as below:

$$\Xi = G\Omega G' = LDL', \quad (5.22)$$

where L is a unique lower triangular matrix with ones in its main diagonal and D is a diagonal matrix with positive entries. We call this decomposition LDL from now onwards. Kim (2010b) extends Yan and Zivot (2007)'s model to a 3-variable model with two common factors, using the same LDL decomposition as above.

Our target in this chapter is to measure price discovery across Brazilian and US market, considering two trading venues in the US. We include exchange rate, in order to split exchange rate shocks and price fundamental's change. We also add common shares for the companies that present enough liquidity for both common and preferred shares in all markets. This leads to a minimum of four-variable model to a maximum of seven-variable model. As we work with high frequency data, the variance of the innovations in its reduced form is much smaller than one. We do not want to force the variance of the structural innovations ($\text{Var}(\eta_t)$) to be equal to one, hence we need a methodology that allows it to be different than the identity matrix, nevertheless still diagonal. Our first choice is to follow Yan and Zivot (2007). Applying LDL decomposition on matrix G , as in (5.17), brings a bias to the estimates in models with more than one cointegrating vector (which is always the case in this chapter). This is illustrated with the Monte Carlo simulations in Section 5.7. We try to correct this problem by constructing matrix G in an alternative way, in particular, on the way to recover the transitory innovations.

Our proposed methodology is to recover the reduced form permanent and transitory innovations using matrix G^* . G^* is built, as in Warne (1993), with $\alpha'\Omega^{-1}$ instead of β' , viz.

$$G^* = [\alpha'_{\perp}, \alpha'\Omega^{-1}]' \quad (5.23)$$

The intuition here is the same as the one when using β' , i.e. to get everything that is transitory and therefore vanishes away. The drawback of β' is that when you have more than one cointegrating vector, it does not work properly. The Monte Carlo exercises show that this change solves the problem of bias. A similar result could be obtained by implementing the Cholesky decomposition as in Gonzalo and Ng (2001).

The second methodological issue that we address is with relation to correlation between common factors. To go from innovations in their reduced form to structural innovations, we need to decompose the covariance matrix. By using the well known Cholesky decomposition, there is the imposition that the variance of the structural innovations are equal to one and most important, that the first common factor can only have impacts from the first structural innovation, the second can only have impacts from the first and the second and so on. Looking at equations (5.1), (5.2) and (5.3), these restrictions mean that λ and π are equal to zero. With the slightly modified version of the Cholesky decomposition proposed by Yan and Zivot (2007), they impose that not only λ and π are equal to zero, but also ρ and κ . This happens because they force the innovations associated with the common factors to have impact equal to one in the long-run, which makes them uncorrelated between each other and hence, equal to the structural innovations.

What we address differently in this chapter is that shocks on common factors might be correlated, hence we need to recover innovations associated to each common factor allowing for this correlation. We construct that by saying that each common factor may respond to three different structural innovations (it could be more than that, but we use the three following equations (5.1), (5.2) and (5.3)). In order to allow the parameters in these equations to be different than zero, we use the spectral decomposition on a normalized covariance matrix. By normalizing the matrix, we impose only one restriction: the variance of the structural permanent and transitory innovations must be equal to the variance of the permanent and transitory innovations in their reduced form. With the normalized matrix, we gain $(k^2 - k)/2$ equations (as we show in Section 5.6), which allows us not to impose the restrictions Cholesky and LDL decomposition impose. In summary, we just implement a decomposition that does not give us any predefined shape on the decomposed matrices, allowing for the parameters of equations (5.1), (5.2) and (5.3) to have any value.

Hence, we decompose the variance matrix of ε_t using the spectral decomposition. With this change, we allow for the variance of η_t to be different than the identify matrix (opposed to Gonzalo and Ng (2001)'s methodology) and at the same time we do not have the need to restrict the long-run impact of permanent innovations to be equal to either one or zero (as it is the case on Yan and Zivot (2007)). This brings a significant benefit, specially when one realizes that the long-run impact can be different than one and zero, as the empirical results prove.

The first step is to normalize Ξ such that $\tilde{\Xi} = \Xi\Theta^{-1}$, with Θ being a diagonal matrix constructed with the diagonal elements of Ξ . In order to

identify the model, we need to impose $\text{Var}(\eta_t) = \Theta$, as shown in Section 5.6. When we implement the spectral decomposition on $\tilde{\Xi}$, we have the following:

$$\tilde{\Xi} = \tilde{S}\tilde{S} \quad (5.24)$$

where \tilde{S} is the squared root of $\tilde{\Xi}$ obtained from an eigenvalue decomposition.

$$\tilde{\Xi} = V\Lambda V^{-1} \Rightarrow \Xi^{1/2} = V\Lambda^{1/2}V^{-1}, \quad (5.25)$$

where the columns of V are the eigenvectors of $\tilde{\Xi}$ and Λ is a diagonal matrix with the corresponding eigenvalues. Hence, we can recover Ξ just by multiplying back Θ .

$$\Xi = \tilde{\Xi}\Theta = \tilde{S}\tilde{S}\Theta \quad (5.26)$$

If we prove that $\tilde{S}\tilde{S}\Theta = \tilde{S}\Theta\tilde{S}'$, we can use exactly the same steps defined above to recover η_t . To this purpose:

$$\tilde{S}\Theta\tilde{S}' = \Xi \quad (5.27)$$

$$\text{Var}(\eta_t) = \Theta = \tilde{S}^{-1}\Xi\tilde{S}^{-1'}$$

$$\eta_t = \tilde{S}^{-1}\varepsilon_t.$$

In Section 5.6, we present the proof of $\tilde{S}\tilde{S}\Theta = \tilde{S}\Theta\tilde{S}'$. This methodology can also be applied when one desires to set $\text{Var}(\eta_t) = I$. We then decompose

Ξ , as below:

$$\Xi = G^* \Omega G^* = S S' \quad (5.28)$$

where S is the squared root of Ξ obtained from the eigenvalue decomposition of matrix Ξ as stated in (5.25). We show in Section 5.6 the identification steps for this case.

The benefit of using $\alpha' \Omega^{-1}$ instead of β' is mainly on normalization. The use of β' carries the question on how to normalize the cointegrating vectors. The standard way is to implement the triangular normalization, however this makes the uniqueness of parameters slightly hard to compute, since for each different order, one would also need to change the cointegrating vectors, not keeping the original triangular normalization. With $\alpha' \Omega^{-1}$ there is no question mark here. The construction of G is straightforward from the results of the VEC model, delivering uniqueness on the parameters.

Regarding the use of spectral decomposition, the main contribution is that common factors might have innovations that are correlated, and so, these innovations are not in their structural form yet. Hence, we choose a methodology that allows to retrieve them in their structural form. Our empirical results show that the parameters on equations (5.1), (5.2) and (5.3) are not zero, indicating that innovations associated with the common factors are indeed correlated. In fact, we find that an innovation related to the efficient exchange rate has a long-run (by long-run we mean one trading day) impact on the observed prices higher than the expect arbitrage adjustment. The main benefit is that we allow long-run effects to be different than one or zero. By allowing this, we include in our model long-run

behavior of innovations on common factors derived from other common factors, extracted from their contemporaneous correlation. This permits a dynamic process not only on observed prices, but also on the common factors themselves. We show on the Monte Carlo simulations that this methodology achieves the best results in finite sample for models with one and two common factors. Extension to the case of more common factors can be easily implemented.

5.4 Price discovery for Brazilian cross listed stocks

We have two targets in this section. The first one is related to price discovery analysis, where we are keen on finding what are the roles of each of the markets in the price discovery dynamics. Secondly, we want to check whether the model presented in Section 5.2 is a valid model for a dynamic process of price discovery.

Starting with the latter one, we find a significant difference between instantaneous and long-run effects, averring that γ_i (in equations (5.5) to (5.9)) is different than unit and there is a partial adjustment process given a shock on the permanent innovations. Hence, the model stated in Section 5.2 seems to fit well in terms of partial adjustments. Important to mention that, as in Section 5.2, by long-run we mean hours within a day, since we are dealing with high frequency data.

As the market microstructure model shows, we believe the efficient price does incorporate part of shocks on the permanent innovations associated with both the exchange rate and voting premium. The same is true for the

efficient voting premium and the efficient exchange rate (both being affected by shocks on the permanent innovation associated with the efficient price). These would firstly imply that the parameters λ , κ , π and ρ to be different than zero. Furthermore, if the direction of the cross linkages among the assets are the ones discussed in section 5.2, we should expect the parameters signs to reflect that.

Indeed, we do find the majority of the elements of $D(1)$ being statistically different than zero across all companies. Therefore, the conclusions on that are twofold: in the long-run, a depreciation of the Brazilian currency leads to a decrease on the value of the firm that exceeds the expected arbitrage adjustment. Second, a positive innovation in the efficient voting premium leads to an increase in the asset's value. These are exactly what we infer when analysing the parameters in Section 5.2: $\pi > 0$ and $\rho < 0$ respectively. As we are not able to identify all the parameters in some equations (specially in foreign common shares, where the number of parameters is higher), we do not find values of λ and κ as high as we find for π and ρ . When we are able to identify them, we find λ very close to zero. Although we expected to find a positive κ , we do not find strong evidences on that, finding it to be closer to zero. We are also not able to identify all the γ parameters, however, it is not so much of interest to look at them individually, since the effects of innovations on the efficient prices are given by the combination of parameters, which is what we estimate and analyse below.

Tables 5.4, 5.1, 5.4, 5.6, 5.7 and 5.8 report results for the six companies considered in this chapter. For the first four companies (Gerdau, BR Telecom, Bradesco and Ambev) there are four markets: exchange rate

(Brazilian Reais/USD dollars), preferred shares traded at the Brazilian market, at NYSE and at ARCA. The shares traded at Brazil are quoted in Brazilian Reais (R\$) and the shares traded at the US market are expressed in US dollars. For these four companies, we find two cointegrating vectors (β), which are in the last two columns of their tables. This leads us to two common factors, seen as the efficient exchange rate and the efficient price. Hence, we analyse the first two columns of D_0 and $D(1)$, since these are the ones related to the two permanent innovations. We call these permanent innovations p^e and p^m , permanent innovation on efficient exchange rate and efficient price respectively. D_0 has the instantaneous effect of a permanent innovation, whereas $D(1)$ has the total effect, as defined and computed in Section 5.3.

For the last two companies (Petrobras and Vale) there are seven and five markets respectively: exchange rate, preferred and common shares traded at the Brazilian market and at NYSE. Petrobras has also common and preferred shares at ARCA. For these two companies we find four and two cointegrating vectors (β), respectively. This leads us to three common factors for both companies, seen as the efficient exchange rate, the efficient price and the efficient voting premium. The observed voting premium is the difference between the common and preferred shares. The efficient voting premium is rid of transitory effects, such as differences in liquidity between the two stocks. We analyse the first three columns of D_0 and $D(1)$, since these are the ones related to the three permanent innovations. We call these permanent innovations p^e , p^m and p^v , as in our theoretical model, where they stand for permanent innovations on the efficient exchange rate, efficient price and efficient voting premium, respectively.

In general, we find the US market as the most important for the price discovery process. In particular, ARCA impounds more information than NYSE which can be explained by the fact that ARCA has a smarter router system. This router is able to check among other exchanges if there is a better quote than the one at ARCA. If this is the case, it executes the order at the venue where the best quote is available. This special characteristic seems to give a more important role for ARCA in the price discovery process when compared to NYSE. In addition to that, it is important to give attention to how liquid (in terms of number of trades) the stocks are in each exchange. ARCA has a similar or higher number of trades than NYSE for the majority of stocks, apart from BR Telecom and Ambev. For these two companies, ARCA has half the number of trades than NYSE, which might affect the importance of ARCA. In the long-run, ARCA and NYSE are equally important.

The highest importance of the US market may be explained by the characteristics of investors in the two markets. Brazilian investors do trade at the US market (reasons might include the fact that US is a much bigger market with a higher potential for diversification, exchange rate issues, etc). It is less likely, however, for an US investor to trade at the Brazilian market, when the stock is available in the US. Thus we would have two sources of information in the US market, whereas only one in Brazil.

Since the US market is the most important, one would expect this market to incorporate not only innovations on the unobserved efficient price, but also innovations on the exchange rate. If the US market is faster on getting permanent news regarding the intrinsic value of the company, it also shows to be faster on adjusting the share price given a permanent

innovation on exchange rate.

Regarding the effect of innovations on the efficient exchange rate on prices, there is an instantaneous overshooting. Given a unit shock on the efficient exchange rate (R\$/USD), i.e. a depreciation of the Brazilian currency, there is a higher depreciation instantaneously than in the long-run. We observe this behavior for all stocks. For the BR Telecom case, we find that exchange rate overshooting is significantly smaller than the ones found for the other companies. This might be explained by the fact that we use 300s interval, leading artificial longer period assigned as short-run. Looking at the theoretical model in Section 5.2, this would be explained by parameter γ_1 being higher than the unit. Intuitively, it could be a signal of herd behavior during turbulent periods, specially if one considers that the Brazilian currency devaluated 49% over 90 days during mid July 2008 and beginning of October 2008 ².

The US market is the one that adjusts the price instantaneously, given a change in the exchange rate. In the long-run, we find that a depreciation of the Brazilian currency actually devalues Brazilian assets, since the parameters we find are negative for the Brazilian market and higher then one in absolute value for the US market (same amount as the Brazilian plus one unit, all negative). This comes as a strong finding, since it associates exchange rate shocks to value of assets using high frequency data. We find this result for all stocks, although some appear to have a stronger depreciation than others. Gerdau, Vale and Petrobras range from 0.50 to 0.60, while BR Telecom and Bradesco are in 0.40's and Ambev is the one to have the least depreciation of asset, with 0.25. This goes in line with

²<http://www.oanda.com/currency/historical-rates>

the finding of $\rho < 0$ in our the theoretical model.

Regarding innovations on the efficient price, the US market has the highest parameters for all stocks, but Ambev. When we compare the importance of NYSE and ARCA, ARCA seems to be faster in incorporating news. We do not find this results only for BR Telecom and Ambev, explained by the fact the these two stocks are so traded at ARCA as they ate at NYSE. Ambev seems to be less affected by exchange rate innovations and is the only one where the Brazilian market is more important than the US market.

We now move to the voting premium aspect looking at Petrobras and Vale, since they are the companies presenting enough trades at common and preferred shares that allowed this analysis. Given a shock on the innovation associated to the efficient voting premium, the common stocks suffer an instantaneous overshooting, while the preferred shares have a negative impact, adjusting for this overshooting. In the long-run, preferred and common shares have a positive impact from a shock on the voting premium. Hence, an increase in the voting premium of a company increase the value of its asset. Again, this is the same result as the one in the theoretical model, where π is higher than zero.

5.5 Robustness

To check the validity of our main results, we perform two robustness checks. The first one is with relation to time stamp, checking whether we see a difference in terms of market leadership across different periods of time. The second one checks if the way we aggregate the data may impact the

final results, specially for the stocks where we have to aggregate at a lower frequency. This check is particularly important for the long-run effects, since we do expect to have differences on the short-run effects given distinct frequencies.

5.5.1 Rolling Window

We perform a rolling window exercise as robustness check for our main results on instantaneous effects and long-run impact. We estimate our model considering a smaller sample size, such that each regression accounts for approximately two months. We do not want to have a very small period and not be able to capture the price dynamics, but at the same time, we look for a reasonable number of windows. The shift window size is set to have the number of observations closely resembling two weeks.

Figures 5.1 to 5.6 displays the price dynamics over time. A few conclusions arise: although the majority of measures is considerably stable over time (specially if we consider the bootstrap intervals also graphed), there are changes in market's importance, specially during the second half of 2008 and first half of 2009. We claim this change of behavior comes from uncertainties derived from the 2008/2009 crises, since the stability is recovered by the end of 2009. This period presents changes in the number of cointegrating vectors for Vale and Petrobras. This is the reason why we do not report the impact on prices derived from shocks on innovations associated with the voting premium, given that we would have missing estimates. We claim the change from three to two common factors comes from the elimination of the voting premium as a common factors in certain periods of the data set. As what we find with the same data set in the

previous chapter, there is a clear change in investor's behavior during the crises period regarding voting premium. They point out this may come from the fact that during turbulent periods, financial assets tend to be more correlated and even share a single common factor among them.

5.5.2 Interval Frequency

In Chapter 3, we explain how we deal with non synchronous trading. Depending on the number of observations each share presents, we accommodate the size of the interval to aggregate the series. For instance, if share A has ten trades for each 30-second interval, and share B had ten trades for each 3-minute interval, we can not aggregate them in 30 seconds, since we would be incurring in a high risk of serial correlation for share B. At the same time, we do not want to aggregate at 3 minutes and lose some important information from share A. We need to find a situation in between. We estimate the covariance matrix using the Newey-West estimator as a way to overcome the serial correlation issue. Additionally, we sample the data at different frequencies, checking whether our main conclusion on market leadership change. This is specially important for stocks where we had to sample at a much lower frequency, for instance 240 seconds. The results regarding long-run impacts do not alter with this change, implying that our results are robust to different sampling frequencies.

5.6 Identification Issues

Regarding the identification strategy two issues have to be addressed to retrieve the price discovery measures. The first one refers to the imple-

mentation of the spectral decomposition, instead of the ones previously adopted in the literature: LDL and Cholesky. The second issue relates to the computation of matrix G , since we allow for specifications containing up to three common factors.

To show that the spectral decomposition can replace either the LDL or the Cholesky decompositions, we need to show that it carries the same number of restrictions as both alternative decompositions. To this purpose, we consider two different scenarios: $\text{Var}(\eta_t) = I$ and $\text{Var}(\eta_t) \neq I$. The reason why we consider these two situations is because in the empirical exercises we do allow for the $\text{Var}(\eta_t)$ to be different than an identity matrix. As we are dealing with high frequency data, the variance of the reduced form error terms, Ω , is very small. Hence, if $\text{Var}(\eta_t)$ is set to be equal to an identity matrix, the resultant parameters in the G matrix would be unrealistic small.

Let us start with the simpler case: $\text{Var}(\eta_t) = I$. The matrix $\Xi = S \text{Var}(\eta_t) S'$ has $(K^2 - K)/2 + K$ equations and $K^2 + (K^2 - K)/2 + K$ unknown variables. Hence, in order to completely identify the model, we need to add further K^2 restrictions. These identify S and $\text{Var}(\eta_t)$. The first set of restrictions comes from the assumption governing the variance of permanent and transitory errors. We assume these innovations have unit variance, which adds K restrictions to our model. The second set of restrictions arises from the use of the structural framework, where $\boldsymbol{\eta}_t = (\eta_t^e, \eta_t^m, \eta_t^v, \boldsymbol{\eta}_t^T)'$ is the vector with all innovations on their structural form. This implies that $\text{Var}(\eta_t)$ is a diagonal matrix, i.e, permanent and transitory shocks are uncorrelated. This adds $(K^2 - K)/2$ restrictions. Finally, when the spectral decomposition is applied to a symmetric matrix, it decomposes symmetric

matrices. Hence, as Ξ is a symmetric matrix, likewise is S , adding the final $(K^2 - K)/2$ restrictions needed.

The second scenario addresses the case which $\text{Var}(\eta_t) \neq I$. To show that the spectral decomposition also holds in this case, we need to change the first and third set of identification restrictions from the previous scenario. We replace the first set of restrictions by imposing $\text{Var}(\eta_t) = \Theta$. This yields the same K restrictions we consider in the previous example. The third set of conditions arises by applying the spectral decomposition to a nonsymmetric matrix $\tilde{\Xi}$, implying that the resulting decomposed matrix is no longer a symmetric matrix, adding additional $(K^2 - K)/2$ equations which completely identify the model. We show in Section 5.3 that if we are able to prove that $\tilde{S}\tilde{S}\Theta = \tilde{S}\Theta\tilde{S}'$ holds, we can recover η . To this purpose, we show the proof for a 2×2 matrix. Define Ξ as:

$$\Xi = \begin{pmatrix} a & b \\ b & c \end{pmatrix} \quad (5.29)$$

Define Θ as a diagonal matrix containing the vector $\theta = (a, c)'$ on its diagonal. Hence, we compute $\tilde{\Xi}$ as

$$\tilde{\Xi} = \Xi\Theta^{-1} = \begin{pmatrix} 1 & \frac{b}{c} \\ \frac{b}{a} & 1 \end{pmatrix}. \quad (5.30)$$

Define V as the matrix containing the eigenvectors associated with $\tilde{\Xi}$ and

Λ the diagonal matrix with the eigenvalues of $\tilde{\Xi}$ on its diagonal, such that:

$$V = \begin{pmatrix} -\frac{\sqrt{a}}{\sqrt{c}} & \frac{\sqrt{a}}{\sqrt{c}} \\ 1 & 1 \end{pmatrix} \quad (5.31)$$

$$\Lambda = \begin{pmatrix} \frac{\sqrt{ac}-b}{\sqrt{ac}} & 0 \\ 0 & \frac{b+\sqrt{ac}}{\sqrt{ac}} \end{pmatrix} \quad (5.32)$$

By applying the spectral decomposition, we have that $\tilde{\Xi} = \tilde{S}\tilde{S}$ and $\Xi = \tilde{S}\tilde{S}\Theta$, with \tilde{S} given by:

$$\begin{aligned} \tilde{S} &= V\Lambda^{1/2}V^{-1} = \\ &= \begin{pmatrix} \frac{1}{2} \left[\left(\frac{\sqrt{ac}-b}{\sqrt{ac}} \right)^{1/2} + \left(\frac{b+\sqrt{ac}}{\sqrt{ac}} \right)^{1/2} \right] & \frac{\sqrt{a}}{2\sqrt{c}} \left[\left(\frac{b+\sqrt{ac}}{\sqrt{ac}} \right)^{1/2} - \left(\frac{\sqrt{ac}-b}{\sqrt{ac}} \right)^{1/2} \right] \\ \frac{\sqrt{c}}{2\sqrt{a}} \left[\left(\frac{b+\sqrt{ac}}{\sqrt{ac}} \right)^{1/2} - \left(\frac{\sqrt{ac}-b}{\sqrt{ac}} \right)^{1/2} \right] & \frac{1}{2} \left[\left(\frac{\sqrt{ac}-b}{\sqrt{ac}} \right)^{1/2} + \left(\frac{b+\sqrt{ac}}{\sqrt{ac}} \right)^{1/2} \right] \end{pmatrix} \end{aligned} \quad (5.33)$$

By computing $\tilde{S}\tilde{S}\Theta$ and $\tilde{S}\Theta\tilde{S}'$, as in (5.34), we show that these two quantities are equal to each other, proving that the normalization holds ³.

$$\begin{aligned} \tilde{S}\tilde{S}\Theta &= \tilde{S}\Theta\tilde{S}' = \\ &= \begin{pmatrix} \frac{a}{2} \left[\left(1 - \frac{b}{\sqrt{ac}} \right) + \left(\frac{b}{\sqrt{ac}} + 1 \right) \right] & \frac{\sqrt{ac}}{2} \left[\left(\frac{b}{\sqrt{ac}} + 1 \right) - \left(1 - \frac{b}{\sqrt{ac}} \right) \right] \\ \frac{\sqrt{ac}}{2} \left[\left(\frac{b}{\sqrt{ac}} + 1 \right) - \left(1 - \frac{b}{\sqrt{ac}} \right) \right] & \frac{c}{2} \left[\left(1 - \frac{b}{\sqrt{ac}} \right) + \left(\frac{b}{\sqrt{ac}} + 1 \right) \right] \end{pmatrix} \end{aligned} \quad (5.34)$$

We solve the second identification issue by showing that we can use some of the rows of $\Psi(1)$ in the place of α'_{\perp} , following Yan and Zivot (2007). Using the Johansen's Factorization as in Johansen (1991), the matrix $\Psi(1)$ can be decomposed as:

$$\Psi(1) = \beta_{\perp}(\alpha'_{\perp}\xi(1)\beta_{\perp})^{-1}\alpha'_{\perp} = \Gamma\alpha'_{\perp} \quad (5.35)$$

³A numerical exercise showing that (5.34) holds for matrix with dimensions greater than two is available upon request.

If we multiply both sides by the error term obtained from the reduced form VEC model, we have

$$\Psi(1)\epsilon_t = \Gamma\alpha'_\perp\epsilon_t \quad . \quad (5.36)$$

Matrix G is built in such a way that the right-hand side of (5.36) contains the portion of ε_t related to permanent innovations, $\varepsilon_t^P = (\varepsilon_t^e, \varepsilon_t^m, \varepsilon_t^v)'$, since ϵ_t is multiplied by the upper part of matrix G .

$$\Psi(1)\epsilon_t = \Gamma\varepsilon_t^P \quad (5.37)$$

From (5.37), the long-run impact of changes in ε_t^P on the market prices is given by Γ . Considering the a model that accounts for exchange rate, preferred and common shares traded at both domestic and foreign markets as the one discussed in Section 5.2, we need to add assumptions that allow us to identify α'_\perp using the rows of $\Psi(1)$. This is a modification of the original identification strategy proposed by Gonzalo and Granger (1995), where they assume that permanent innovations present a long-run effect different than zero, whereas transitory shocks vanish away in the long-run. Assuming a simpler model than mine, Kim (2010a) and Yan and Zivot (2010) impose long-run restrictions on the permanent innovations in their reduced form to justify the use of common rows in $\Psi(1)$ to identify α'_\perp . Our identification strategy follows along these lines, but we recover matrix Γ in (5.36) using the parameters that drive the common factors dynamics. This covers the case where $\Psi(1)$ does not have clear common rows. To this

purpose, from (5.1), (5.2) and (5.3), we construct a matrix Φ such that:

$$\Phi = \begin{pmatrix} 1 & \lambda & 0 \\ \rho & 1 & \pi \\ 0 & \kappa & 1 \end{pmatrix} \quad (5.38)$$

Equation (5.11) gives the long-run dynamics of prices as function of the permanent and transitory innovations on their structured form. Define $D_P(1)$ as the sub-matrix containing all rows and the first three columns of the $D(1)$ matrix. Hence, we want to impose restrictions on Γ in the right-hand side of (5.37) such that the long-run dynamics depicted in (5.11) holds. To this purpose, it is sufficient to find Γ that makes $D_P(1) \eta_t = \Gamma \varepsilon_t^P$, provided that $\varepsilon_t^P = \Phi \eta_t^P$ holds.

$$\begin{aligned} \Gamma \varepsilon_t^P &= D_P(1) \eta_t^P \\ \Gamma \Phi \eta_t &= D_P(1) \eta_t^P \\ \Gamma \Phi &= D_P(1) \\ \Gamma &= D_P(1) \Phi^{-1} \end{aligned} \quad (5.39)$$

Hence, the left-hand side of (5.39) resumes to:

$$\begin{aligned}
\Gamma &= \begin{pmatrix} 1 & \lambda & 0 \\ \rho & 1 & \pi \\ \rho & \kappa + 1 & \pi + 1 \\ \rho - 1 & 1 - \lambda & \pi \\ \rho - 1 & \kappa - \lambda + 1 & \pi + 1 \end{pmatrix} \times \\
&\quad \times \begin{pmatrix} \frac{1-\kappa\pi}{-\kappa\pi-\lambda\rho+1} & -\frac{\lambda}{-\kappa\pi-\lambda\rho+1} & \frac{\lambda\pi}{-\kappa\pi-\lambda\rho+1} \\ -\frac{\rho}{-\kappa\pi-\lambda\rho+1} & \frac{1}{-\kappa\pi-\lambda\rho+1} & -\frac{\pi}{-\kappa\pi-\lambda\rho+1} \\ \frac{\kappa\rho}{-\kappa\pi-\lambda\rho+1} & -\frac{\kappa}{-\kappa\pi-\lambda\rho+1} & \frac{1-\lambda\rho}{-\kappa\pi-\lambda\rho+1} \end{pmatrix} \\
\Gamma &= \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 1 & 1 \\ -1 & 1 & 0 \\ -1 & 1 & 1 \end{pmatrix} \tag{5.40}
\end{aligned}$$

Combining (5.37) with (5.40), we have that the first two rows of $\Psi(1)$ can be used in place of the first two rows of α'_\perp and the third minus the second row of $\Psi(1)$ as the third row of α'_\perp . These imply that the reduced form innovations associated with the efficient price do not have a permanent impact on the exchange rate in the long-run. This is not harmful in our analysis for two reasons: first, there is no reason to imagine that an innovation on the efficient price should have an effect on exchange rate, apart from correlation aspects governing the structural innovations associated to the common factors. Second, if there are such correlation among the structural errors, this still can be captured by the model, since the restriction

is on ε_t^m and not on η_t^m . Moreover, Γ imposes that changes in the reduced innovation associated with the exchange rate affects only the foreign market. Again, this is not harmful, since both restrictions are constructed in terms of ε^e .

5.7 Simulations

This section illustrates our proposed estimation methodology by comparing it with the alternative frameworks available in the literature. We focus our analysis on computing the instantaneous and long-run measures for price discovery. As pointed out in Section 5.3, we propose two changes in the methodology. The first one refers to the computation of matrix G , using $\alpha'\Omega^{-1}$ instead of β' . The second one adopts the spectral decomposition rather than the LDL or Cholesky decompositions.

The model we use here is a simplified version of the one presented in Section 5.2. We work with two common factors, but the extension to the case with more common factors is straightforward. We also assume that the parameters λ , ρ , π and κ are all equal to zero, which simplifies our results. Given these restrictions, the elements of D_0 are the parameters giving the partial adjustment between efficient and observed prices. We additionally assume that the efficient exchange rate is an observed process. Therefore, the data generation process is given by:

$$e_t = e_{t-1} + \eta_t^e \tag{5.41}$$

$$m_t = m_{t-1} + \eta_t^m$$

where e_t is the efficient exchange rate and m_t is the asset efficient price. The

structural innovations η_t^e and η_t^m are random normal processes generated with a diagonal covariance matrix. The transitory innovations η_t^T are also normally distributed. The observed prices are given by:

$$\Delta y_{2,t} = \gamma_2 (m_t - y_{2,t-1}) + b_2 \eta_t^T \quad (5.42)$$

$$\Delta y_{3,t} = \gamma_3 (m_t - y_{3,t-1}) + b_3 \eta_t^T$$

$$\Delta y_{4,t}^* = \gamma_4 (m_t - y_{4,t-1}) - \dot{\gamma}_4 (e_t - e_{t-1}) + b_4 \eta_t^T$$

$$\Delta y_{5,t}^* = \gamma_5 (m_t - y_{5,t-1}) - \dot{\gamma}_5 (e_t - e_{t-1}) + b_5 \eta_t^T$$

where $y_{2,t}$ and $y_{3,t}$ are transactions prices observed in the domestic market, whereas $y_{4,t}^*$ and $y_{5,t}^*$ are prices observed in the foreign market expressed in foreign currency. The 1×3 vector b_i has the parameters accompanying the transitory innovations.

$$D_0 = \begin{pmatrix} 1 & 0 & 0 & 0 & 0 \\ 0 & \gamma_2 & b_{1i} & b_{2i} & b_{3i} \\ 0 & \gamma_3 & b_{1i} & b_{2i} & b_{3i} \\ \dot{\gamma}_4 & \gamma_2 & b_{1i} & b_{2i} & b_{3i} \\ \dot{\gamma}_5 & \gamma_5 & b_{1i} & b_{2i} & b_{3i} \end{pmatrix} \quad (5.43)$$

Table 5.9 reports results based on the four different comparisons. Firstly, we want to measure the benefit of computing D_0 using the matrix G constructed with $\alpha' \Omega^{-1}$. Therefore, we compare \tilde{D}_0 versus \dot{D}_0 , where \dot{D}_0 stands for D_0 computed using $\alpha' \Omega^{-1}$ and decomposed with LDL, whereas \tilde{D}_0 stands for D_0 calculated with the matrix G computed using β' and LDL decomposition. The second comparison assesses the benefit of using only the spectral decomposition. Hence, we compute two estimates of D_0 : the

first one uses the $\alpha'\Omega^{-1}$ expression in G and the spectral decomposition (denoted as \hat{D}_0), whereas the second measure uses $\alpha'\Omega^{-1}$ and the LDL decomposition (denoted as \dot{D}_0). The third comparison addresses the benefits of combining our two methodological suggestions. We compute D_0 using both the $\alpha'\Omega^{-1}$ expression and the spectral decomposition (denoted as \hat{D}_0) and we denote \tilde{D}_0 as the estimates computed using with β' and the LDL decomposition. Finally, we also want to compare \hat{D}_0 with the methodology suggested by Gonzalo and Ng (2001). We denote it as \bar{D}_0 and we compute it using β' and the Cholesky decomposition.

We report results in terms of the mean, relative mean squared errors (RelMSE) and relative root mean squared error (RelRMSE). We display the ratio of the D_0 measures to indicate the way the relative measures are computed. For instance, \hat{D}_0/\bar{D}_0 implies that the relative measures are computed having \bar{D}_0 in the denominator and \hat{D}_0 in the numerator. Thus, relative measures smaller than one indicates that the \hat{D}_0 outperforms \bar{D}_0 .

The results show that \tilde{D}_0 is biased for systems with more than one cointegrating vector (we did compute \tilde{D}_0 for a smaller system with only one cointegrating vector and the bias is eliminated). By inserting $\alpha'\Omega^{-1}$ we are able to eliminate all the bias, and we could even continue to use LDL decomposition, as we see on the results considering \dot{D}_0 . Hence, \dot{D}_0 , \bar{D}_0 and \hat{D}_0 are not biased. By analyzing the relative measures, we show that \hat{D}_0 presents massive gains when compared to the \tilde{D}_0 measures. Similar results are obtained when \dot{D}_0 is compared \tilde{D}_0 , indicating that by using $\alpha'\Omega^{-1}$ instead of β' , we are able to improve considerably our price discovery estimates. In summary, our proposed measure outperforms all competitors.

5.8 Conclusion

We investigate the price discovery for cross listed Brazilian companies in this chapter. We are interested on measuring how fast permanent innovations are impounded by the different platforms, as well as what markets are the most important on incorporating this new information.

We present a simple market microstructure model, that guides the understanding of the empirical results. Our model allows the observed prices to depend on three different common factors: the efficient exchange rate, the efficient asset price and efficient voting premium. Moreover, we allow the common factors to be contemporaneously correlated, yielding the necessary conditions for cross linkages among the common factors. We provide short-run and long-run solutions as function of the structural parameters, as well as price discovery dynamic measures.

We propose an alternative methodology to measure instantaneous effects of permanent shocks on prices in spirit of Yan and Zivot (2007). Our methodology does not present any issues regarding normalization of the cointegrating vectors. It is order invariant and works properly even for a large number of variables and cointegrating vectors. By using the structural framework, we are able to assess whether a permanent shock on the exchange rate changes the company's value more than the expected arbitrage adjustment. This is an interesting point which so far has not been analyzed using the price discovery framework. We find through a Monte Carlo exercise that our measure presents better performance in finite sample, when compared to all competitors.

On the empirical results, we find that the trading platform ARCA is the most efficient market in incorporating shocks instantaneously. US market is

the one that adjust for exchange rate shocks. We also observe that liquidity plays an important role on how fast markets impound new information.

Finally, the theoretical model proposed lead to interesting results regarding the price process. We find that in real terms, Brazilian companies lose value in the “long-run”, following a depreciation of the Brazilian currency and a positive innovation in the efficient voting premium leads to an increase in the asset’s value.

Table 5.1
Price Discovery BR Telecom 300''

Exchange rate (ExRate) is expressed in R\$ per US dollars. Permanent shocks are denoted as η_t^e and η_t^m , where the former one is related to the efficient exchange rate, whereas the latter one refers to the efficient price of the underlying security. We determine lag length in the VEC model through Schwarz criteria. BR Telecom prices are sampled at 300 seconds frequency ($T = 35, 229$). The bootstrap standard errors are in the parenthesis.

		Inst. Effect		Total		Fastness		β
		η_t^e	η_t^m	η_t^e	η_t^m	η_t^e	η_t^m	
	<i>BRLUSD</i>	1.07 (0.018)	-0.06 (0.008)	0.98 (0.003)	-0.08 (0.007)	1.08 (0.017)	0.76 (0.093)	0 -1
BR	<i>BRT04</i>	0.21 (0.074)	0.89 (0.031)	-0.41 (0.039)	0.98 (0.003)	0.52 (0.185)	0.91 (0.031)	0 1
NYSE	<i>BTM_n</i>	-0.88 (0.087)	1.00 (0.035)	-1.40 (0.036)	1.07 (0.005)	0.63 (0.057)	0.93 (0.033)	1 0
ARCA	<i>BTM_p</i>	-0.83 (0.082)	0.96 (0.034)	-1.39 (0.036)	1.07 (0.005)	0.60 (0.054)	0.89 (0.031)	-1 -1

Table 5.2
Price Discovery BR Telecom 240”

Exchange rate (ExRate) is expressed in R\$ per US dollars. Permanent shocks are denoted as η_t^e and η_t^m , where the former one is related to the efficient exchange rate, whereas the latter one refers to the efficient price of the underlying security. We determine lag length in the VEC model through Schwarz criteria. BR Telecom prices are sampled at 240 seconds frequency ($T = 44, 103$). The bootstrap standard errors are in the parenthesis.

	Inst. Effect		Total		Fastness	
	η_t^e	η_t^m	η_t^e	η_t^m	η_t^e	η_t^m
<i>BRLUSD</i>	1.08 (0.016)	-0.06 (0.008)	0.98 (0.002)	-0.08 (0.007)	1.10 (0.016)	0.72 (0.085)
BR	0.19 (0.066)	0.94 (0.03)	-0.41 (0.034)	0.98 (0.002)	0.46 (0.168)	0.95 (0.03)
NYSE	-0.79 (0.083)	0.98 (0.034)	-1.39 (0.032)	1.07 (0.005)	0.56 (0.055)	0.92 (0.032)
ARCA	-0.77 (0.079)	0.96 (0.032)	-1.39 (0.032)	1.07 (0.005)	0.55 (0.053)	0.89 (0.029)

Table 5.3
Price Discovery Bradesco 30”

Exchange rate (ExRate) is expressed in R\$ per US dollars. Permanent shocks are denoted as η_t^e and η_t^m , where the former one is related to the efficient exchange rate, whereas the latter one refers to the efficient price of the underlying security. We determine lag length in the VEC model through Schwarz criteria. Bradesco prices are sampled at 30 seconds frequency ($T = 352, 183$). The bootstrap standard errors are in the parenthesis.

		Inst. Effect		Total		Fastness		β
		η_t^e	η_t^m	η_t^e	η_t^m	η_t^e	η_t^m	
BR	<i>BRLUSD</i>	1.30 (0.019)	0.01 (0.009)	0.98 (0.003)	-0.10 (0.007)	1.33 (0.016)	0.12 (0.093)	0 -1
	<i>BDBC4</i>	0.31 (0.032)	0.78 (0.028)	-0.47 (0.024)	0.98 (0.003)	0.65 (0.068)	0.80 (0.027)	0 1
NYSE	<i>BBD_n</i>	-1.10 (0.061)	0.81 (0.025)	-1.45 (0.022)	1.08 (0.004)	0.75 (0.037)	0.75 (0.022)	1 0
	<i>BBD_p</i>	-1.29 (0.051)	0.97 (0.02)	-1.45 (0.022)	1.08 (0.004)	0.88 (0.025)	0.90 (0.017)	-1 -1

Table 5.4
Price Discovery Gerdau 30’’

Exchange rate (ExRate) is expressed in R\$ per US dollars. Permanent shocks are denoted as η_t^e and η_t^m , where the former one is related to the efficient exchange rate, whereas the latter one refers to the efficient price of the underlying security. We determine lag length in the VEC model through Schwarz criteria. Gerdau prices are sampled at 30 seconds frequency ($T = 352,159$). The bootstrap standard errors are in the parenthesis.

		Inst. Effect		Total		Fastness			
		η_t^e	η_t^m	η_t^e	η_t^m	η_t^e	η_t^m	η_t^e	β
	<i>BRLUSD</i>	1.34 (0.018)	0.00 (0.007)	0.97 (0.003)	-0.09 (0.005)	1.38 (0.016)	0.00 (0.048)	0	-1
BR	<i>GGBR4</i>	0.19 (0.047)	0.84 (0.029)	-0.62 (0.027)	0.97 (0.003)	0.32 (0.073)	0.87 (0.029)	0	1
NYSE	<i>GBR_n</i>	-0.82 (0.094)	0.74 (0.033)	-1.58 (0.025)	1.07 (0.003)	0.52 (0.055)	0.69 (0.031)	1	0
ARCA	<i>GBR_p</i>	-1.34 (0.087)	0.90 (0.029)	-1.58 (0.025)	1.07 (0.003)	0.85 (0.047)	0.84 (0.026)	-1	-1

Table 5.5
Price Discovery Gerdau 60’’

Exchange rate (ExRate) is expressed in R\$ per US dollars. Permanent shocks are denoted as η_t^e and η_t^m , where the former one is related to the efficient exchange rate, whereas the latter one refers to the efficient price of the underlying security. We determine lag length in the VEC model through Schwarz criteria. Gerdau prices are sampled at 30 and 60 seconds frequency ($T = 352, 159$ and $T = 176, 313$). The bootstrap standard errors are in the parenthesis.

		Inst. Effect		Total		Fastness		β
		η_t^e	η_t^m	η_t^e	η_t^m	η_t^e	η_t^m	
BR	<i>BRLUSD</i>	1.20 (0.016)	-0.02 (0.007)	0.97 (0.003)	-0.09 (0.006)	1.24 (0.014)	0.19 (0.066)	0 -1
	<i>GGBR4</i>	0.15 (0.044)	0.79 (0.025)	-0.63 (0.031)	0.97 (0.003)	0.24 (0.069)	0.81 (0.025)	0 1
NYSE	<i>GBR_n</i>	-1.05 (0.099)	0.84 (0.029)	-1.59 (0.029)	1.07 (0.003)	0.66 (0.057)	0.79 (0.026)	1 0
ARCA	<i>GBR_p</i>	-1.48 (0.081)	0.96 (0.024)	-1.60 (0.029)	1.07 (0.003)	0.93 (0.041)	0.90 (0.021)	-1 -1

Table 5.6
Price Discovery Ambev 90’’

Exchange rate (ExRate) is expressed in R\$ per US dollars. Permanent shocks are denoted as η_t^e and η_t^m , where the former one is related to the efficient exchange rate, whereas the latter one refers to the efficient price of the underlying security. We determine lag length in the VEC model through Schwarz criteria. Ambev prices are sampled at 90 seconds frequency ($T = 117,087$). The bootstrap standard errors are in the parenthesis.

		Inst. Effect		Total		Fastness			
		η_t^e	η_t^m	η_t^e	η_t^m	η_t^e	η_t^m	η_t^e	β
BR	<i>BRLUSD</i>	1.14 (0.017)	-0.03 (0.011)	0.99 (0.002)	-0.09 (0.008)	1.15 (0.017)	0.29 (0.107)	0	-1
	<i>AMBEV4</i>	0.17 (0.034)	0.95 (0.032)	-0.26 (0.022)	0.99 (0.002)	0.66 (0.134)	0.97 (0.032)	0	1
NYSE	<i>ABV_n</i>	-0.76 (0.07)	0.85 (0.036)	-1.25 (0.02)	1.09 (0.007)	0.61 (0.054)	0.78 (0.033)	1	0
ARCA	<i>ABV_p</i>	-0.91 (0.072)	0.96 (0.038)	-1.25 (0.02)	1.09 (0.007)	0.72 (0.053)	0.88 (0.034)	-1	-1

Table 5.7
Price Discovery Petrobras 30’’

Exchange rate (ExRate) is expressed in R\$ per US dollars. Permanent shocks are denoted as η_t^e , η_t^m and η_t^v , related to the efficient exchange rate, the efficient price of the underlying security and efficient voting premium respectively. Number 3 stands for common shares, 4 and subscript a for preferred ones. We determine lag length in the VEC model through Schwarz criteria. Petrobras prices are sampled at 30 seconds frequency ($T = 352,676$).

		Inst. Effect			Total			Fastness			
		η_t^e	η_t^m	η_t^v	η_t^e	η_t^m	η_t^v	η_t^e	η_t^m	η_t^v	β
	<i>BRLUSD</i>	1.22 (0.029)	0.05 (0.014)	0.20 (0.043)	0.97 (0.004)	-0.11 (0.008)	0.03 (0.029)	1.26 (0.027)	0.46 (0.157)	7.27 (253.652)	0 0 0 1
BR	<i>PETR4</i>	-0.18 (0.049)	0.86 (0.023)	-0.41 (0.09)	-0.51 (0.033)	0.96 (0.003)	0.54 (0.063)	0.36 (0.087)	0.89 (0.023)	0.76 (0.245)	0 0 1 0
BR	<i>PETR3</i>	0.15 (0.031)	0.46 (0.025)	1.45 (0.069)	-0.50 (0.032)	0.99 (0.005)	1.53 (0.061)	0.30 (0.06)	0.47 (0.024)	0.95 (0.064)	0 0 -1 -1
NYSE	<i>PBRa_N</i>	-1.14 (0.071)	0.90 (0.025)	-0.65 (0.094)	-1.48 (0.03)	1.07 (0.006)	0.49 (0.087)	0.77 (0.046)	0.83 (0.022)	1.34 (0.46)	0 1 -1 0
NYSE	<i>PBR_N</i>	-1.23 (0.062)	0.95 (0.029)	1.92 (0.08)	-1.47 (0.029)	1.10 (0.005)	1.49 (0.086)	0.83 (0.043)	0.86 (0.027)	1.29 (0.1)	1 0 1 1
ARCA	<i>PBRa_P</i>	-1.46 (0.034)	1.00 (0.012)	-0.49 (0.086)	-1.48 (0.03)	1.07 (0.006)	0.49 (0.087)	0.99 (0.017)	0.93 (0.01)	1.01 (0.403)	0 -1 0 0
ARCA	<i>PBR_P</i>	-1.48 (0.036)	1.04 (0.014)	1.79 (0.105)	-1.47 (0.029)	1.10 (0.005)	1.49 (0.086)	1.00 (0.017)	0.95 (0.012)	1.20 (0.033)	-1 0 0 0

Table 5.8
Price Discovery Vale 30''

Exchange rate (ExRate) is expressed in R\$ per US dollars. Permanent shocks are denoted η_t^e , η_t^m and η_t^v , related to the efficient exchange rate, the efficient price of the underlying security and efficient voting premium respectively. Number 3 stands for common shares, 5 and subscript p accounts for preferred ones. We determine lag length in the VEC model through Schwarz criteria. Vale prices are sampled at 30 seconds frequency ($T = 352,344$).

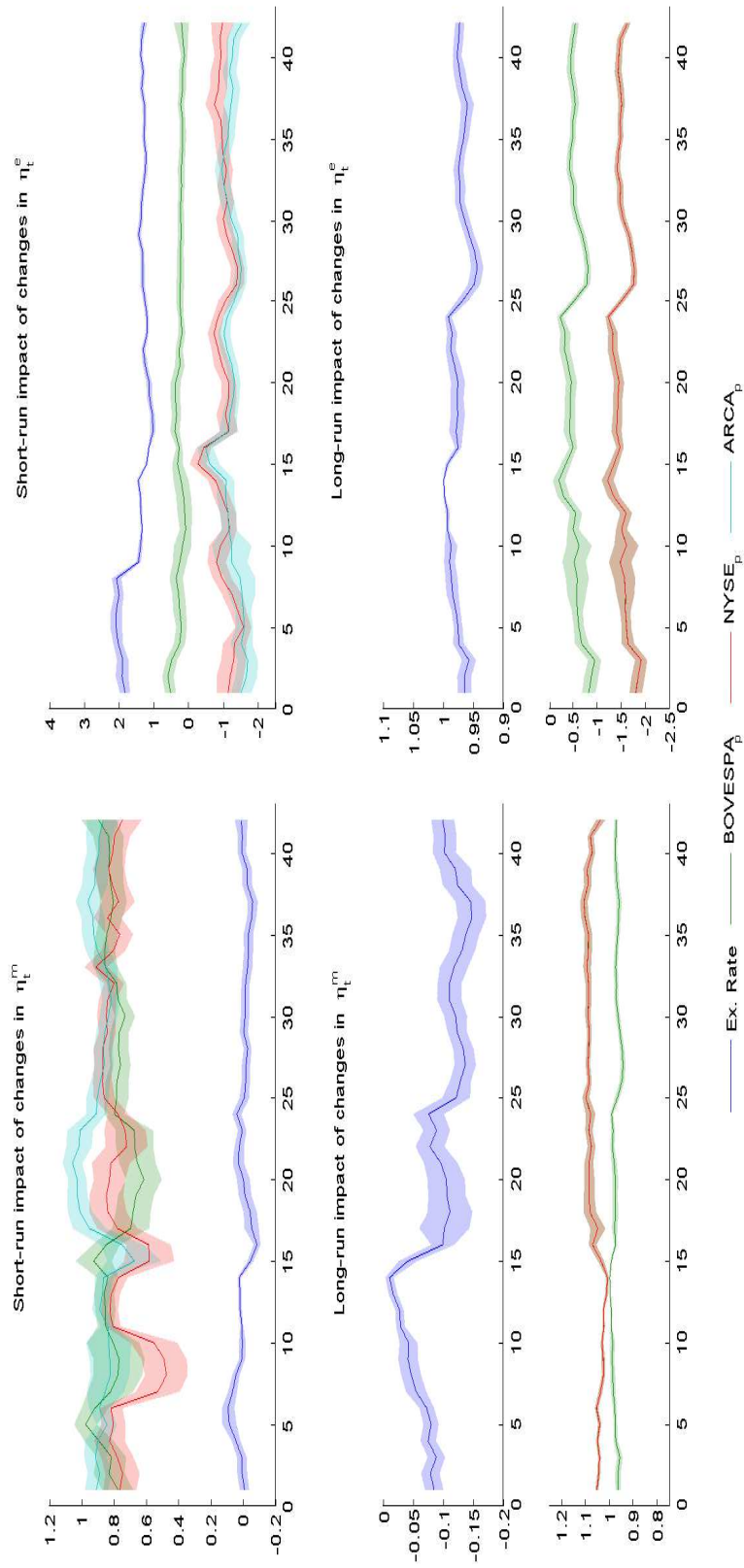
		Inst. Effect			Total			Fastness				
		η_t^e	η_t^m	η_t^v	η_t^e	η_t^m	η_t^v	η_t^e	η_t^m	η_t^v		β
	<i>BRLUSD</i>	1.16 (0.02)	-0.05 (0.012)	0.03 (0.034)	0.96 (0.004)	-0.12 (0.007)	-0.03 (0.017)	1.20 (0.019)	0.43 (0.084)	0.98 (14.026)	1	-1
BR	<i>VALE5</i>	-0.47 (0.06)	0.99 (0.019)	0.18 (0.079)	-0.57 (0.044)	0.95 (0.005)	0.51 (0.042)	0.82 (0.069)	1.04 (0.019)	0.34 (0.143)	0	1
BR	<i>VALE3</i>	-0.53 (0.069)	0.93 (0.025)	2.25 (0.09)	-0.58 (0.046)	0.99 (0.006)	1.50 (0.041)	0.91 (0.082)	0.94 (0.025)	1.50 (0.06)	-1	0
NYSE	<i>RIOp_N</i>	-1.58 (0.047)	1.12 (0.011)	0.17 (0.063)	-1.54 (0.04)	1.08 (0.005)	0.52 (0.051)	1.02 (0.016)	1.04 (0.009)	0.34 (0.106)	0	-1
NYSE	<i>RIO_N</i>	-1.61 (0.05)	1.14 (0.016)	2.31 (0.07)	-1.55 (0.042)	1.11 (0.005)	1.53 (0.049)	1.04 (0.019)	1.03 (0.014)	1.51 (0.046)	1	0

Table 5.9
Monte Carlo Simulations

Results are expressed in terms of Relative Mean Squared Error (RMSE) and Relative Root Mean Squared Error (RRMSE). Sample size and replication number are fixed at 10,000 and 1,000 respectively. The variable $d_0^{i,j}$ denotes the ij^{th} element of the D_0 matrix.

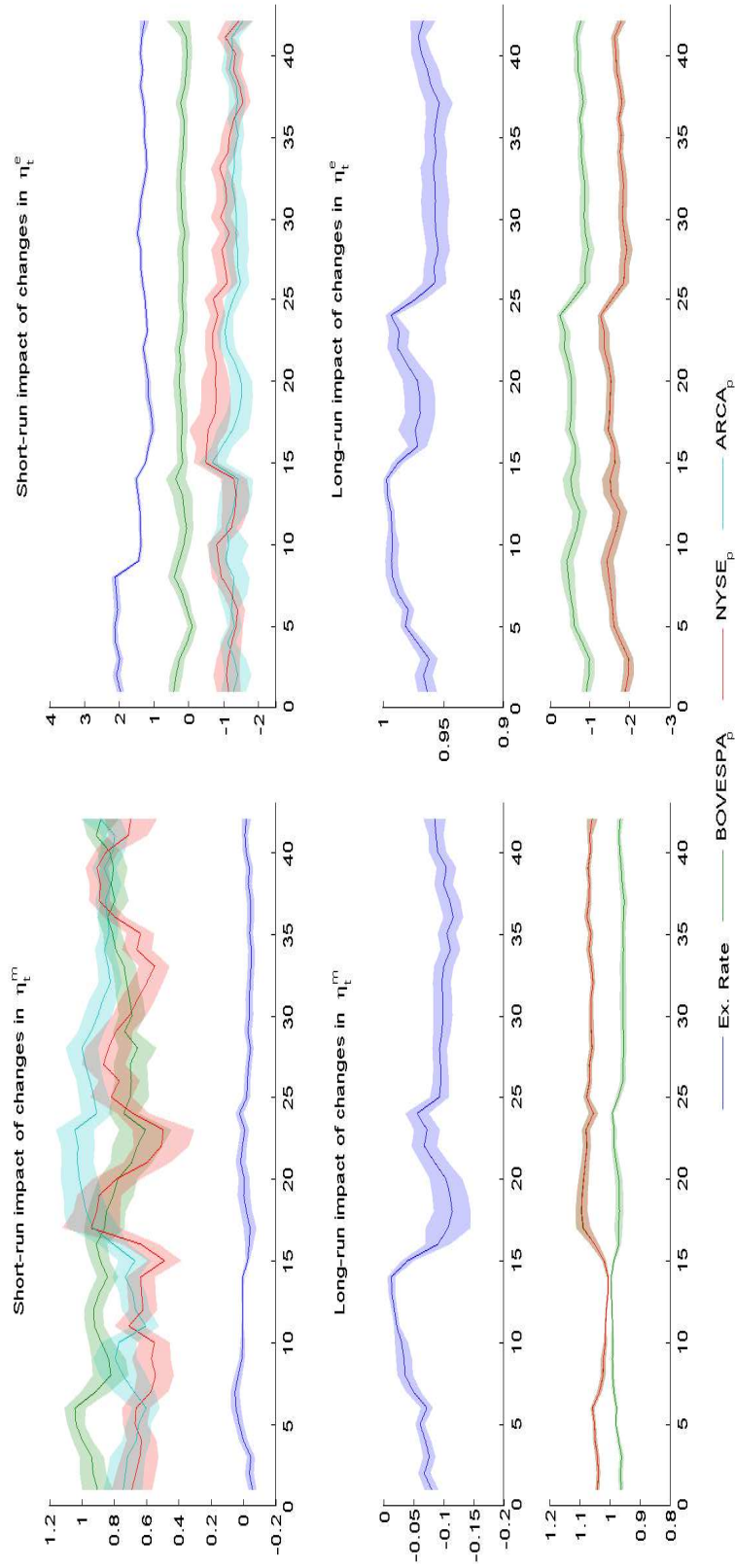
True value	Mean			RMSE			RRMSE		
	β_{LDL}	α_{LDL}	β_C	$\frac{\alpha_S}{\alpha_{LDL}}$	$\frac{\alpha_S}{\beta_{LDL}}$	$\frac{\alpha_S}{\beta_C}$	$\frac{\alpha_S}{\alpha_{LDL}}$	$\frac{\alpha_S}{\beta_{LDL}}$	$\frac{\alpha_S}{\beta_C}$
D_0	\tilde{D}_0	\dot{D}_0	\bar{D}_0	\hat{D}_0/\tilde{D}_0	\hat{D}_0/\dot{D}_0	\hat{D}_0/\bar{D}_0	\hat{D}_0/\tilde{D}_0	\hat{D}_0/\dot{D}_0	\hat{D}_0/\bar{D}_0
$d_0^{11} = 1.0$	1.03	0.99	0.99	0.99	1.00	0.92	0.22	1.00	0.22
$d_0^{21} = 0.0$	0.46	0.00	0.00	0.00	0.80	0.65	0.03	0.64	0.02
$d_0^{31} = 0.0$	0.30	0.00	0.00	0.00	0.97	0.93	0.05	0.95	0.05
$d_0^{41} = 0.2$	1.15	0.20	0.20	0.20	0.99	0.97	0.01	0.99	0.01
$d_0^{51} = 0.5$	0.95	0.50	0.50	0.50	0.95	0.87	0.05	0.91	0.05
$d_0^{12} = 0.0$	0.01	0.00	0.00	0.00	0.70	0.51	0.45	0.50	0.32
$d_0^{22} = 0.8$	0.87	0.79	0.79	0.79	1.00	0.99	0.09	1.00	0.09
$d_0^{32} = 0.4$	0.53	0.40	0.40	0.40	1.00	1.00	0.13	1.00	0.13
$d_0^{42} = 0.2$	0.45	0.20	0.20	0.20	0.96	0.93	0.04	0.93	0.04
$d_0^{52} = 0.5$	0.54	0.50	0.49	0.50	0.93	0.86	0.10	0.87	0.10

Figure 5.1
Bradesco



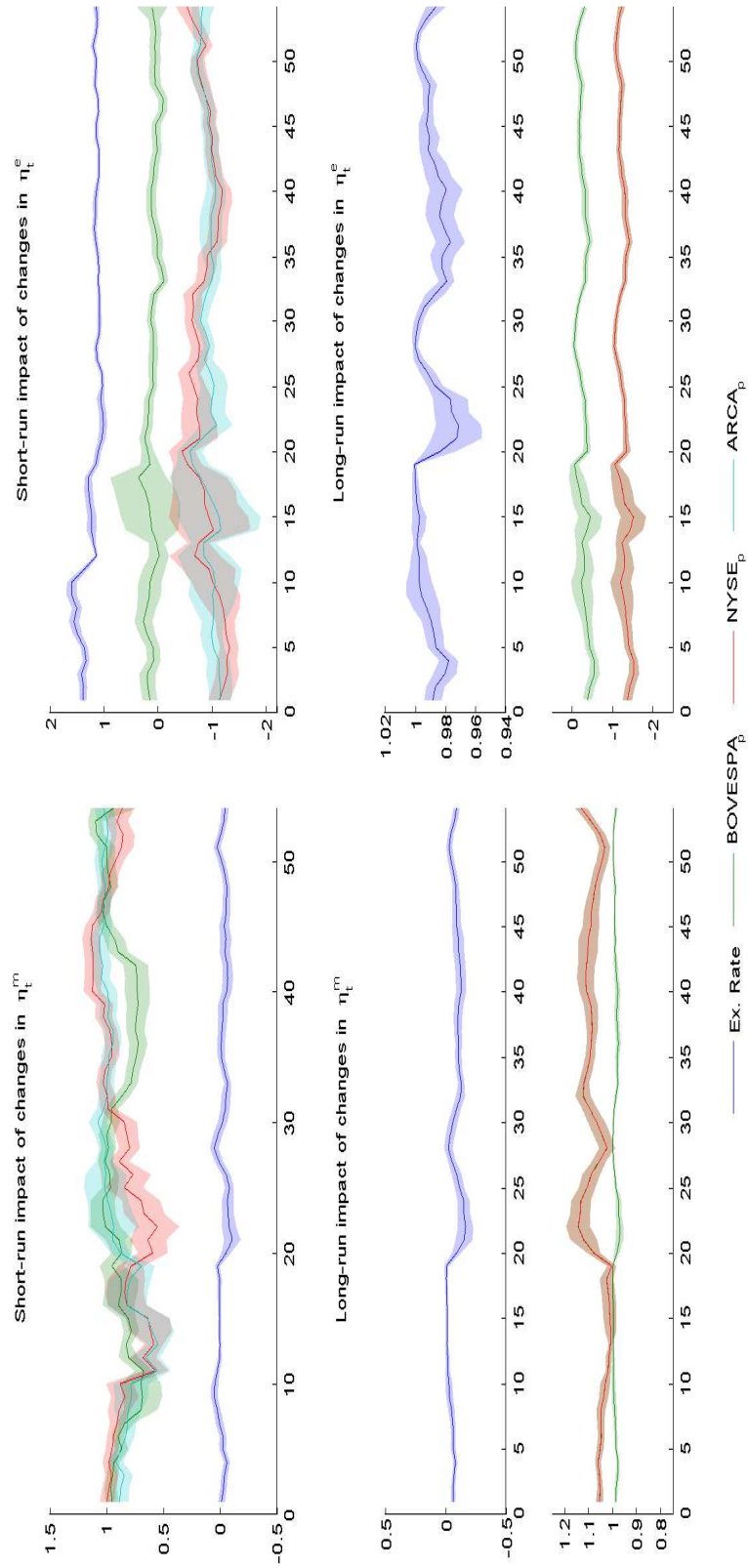
Displays rolling window estimates of the short and long-run structural polynomials, $D(0)$ and $D(1)$ respectively. Sample size and shift window size are fixed equal to 30,547 and 7,637 observations respectively, resulting in 42 windows. Prices are aggregated at 30 seconds. $BOVESPA_p$ accounts for preferred shares traded at BOVESPA (Brazil), $NYSE_p$ for the ADR's on preferred shares traded at NYSE and $ARCA_p$ for preferred shares traded at ARCA. Exchange rate (ExRate) is expressed in R\$ per US dollars. 95% confidence intervals are obtained using the bootstrap standard errors.

Figure 5.2
Gerdau



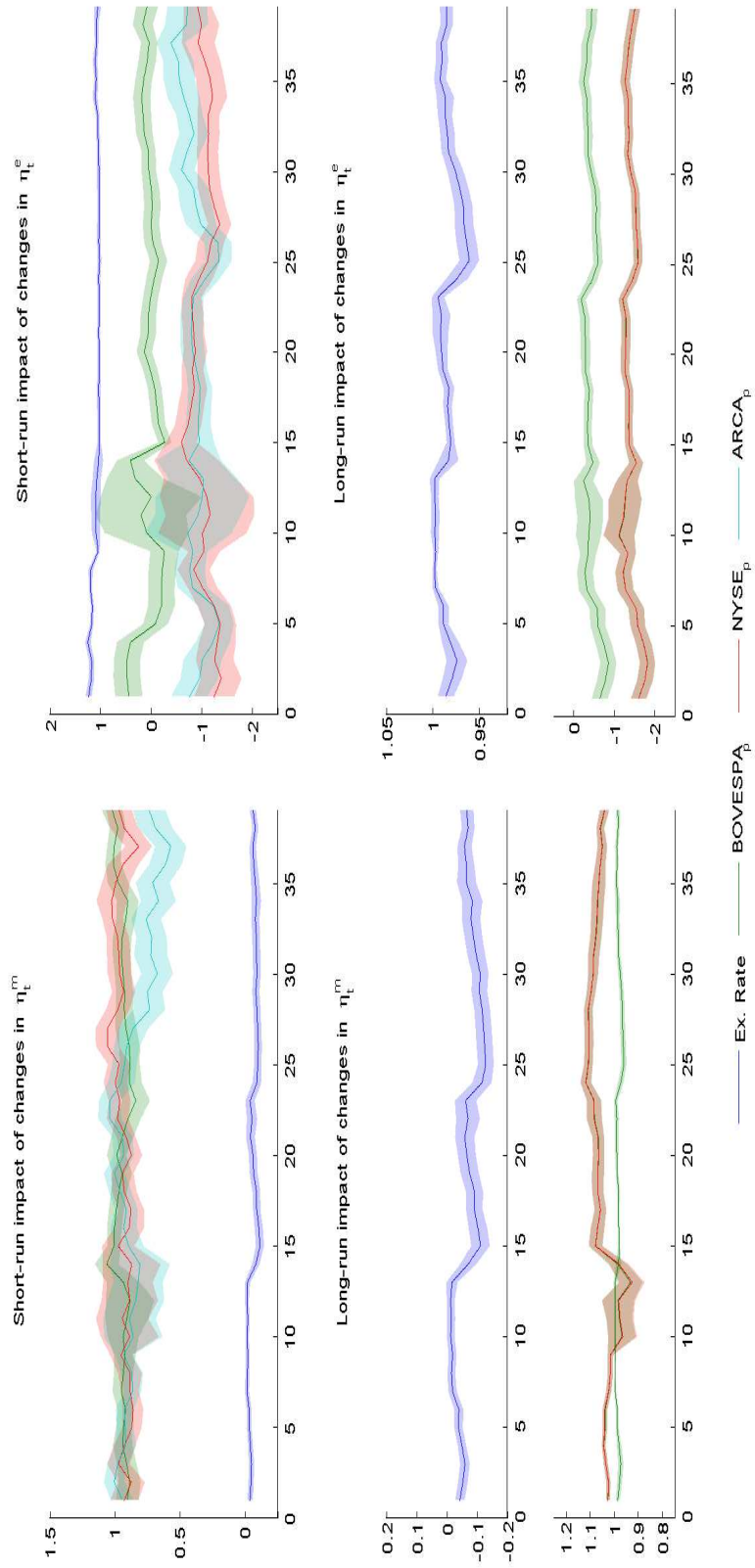
Displays rolling window estimates of the short and long-run structural polynomials, $D(0)$ and $D(1)$ respectively. Sample size and window size are fixed equal to 30,547 and 7,637 observations respectively, resulting in 42 windows. Prices are aggregated at 30 seconds. $BOVESPA_p$ accounts for preferred shares traded at BOVESPA (Brazil), $NYSE_p$ for the ADR's on preferred shares traded at NYSE and $ARCA_p$ for preferred shares traded at ARCA. Exchange rate (ExRate) is expressed in R\$ per US dollars. 95% confidence intervals are obtained using the bootstrap standard errors.

Figure 5.3
Ambev



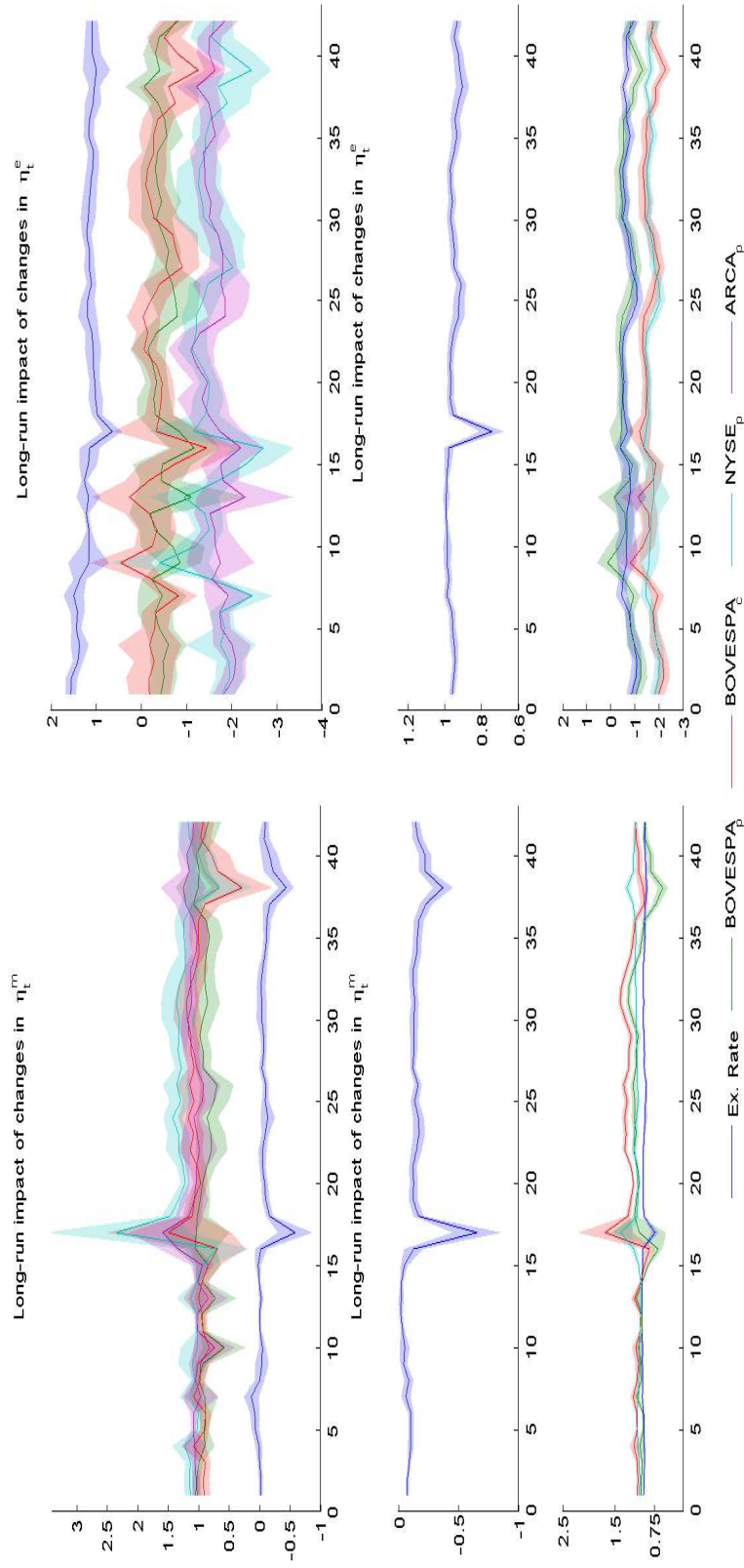
Displays rolling window estimates of the short and long-run structural polynomials, $D(0)$ and $D(1)$ respectively. Sample size and window size are fixed equal to 10,000 and 2,000 observations respectively, resulting in 54 windows. Prices are aggregated at 90 seconds. $BOVESPA_p$ accounts for preferred shares traded at BOVESPA (Brazil), $NYSE_p$ for the ADR's on preferred shares traded at NYSE and $ARCA_p$ for preferred shares traded at ARCA. Exchange rate (ExRate) is expressed in R\$ per US dollars. 95% confidence intervals are obtained using the bootstrap standard errors.

Figure 5.4
BR Telecom



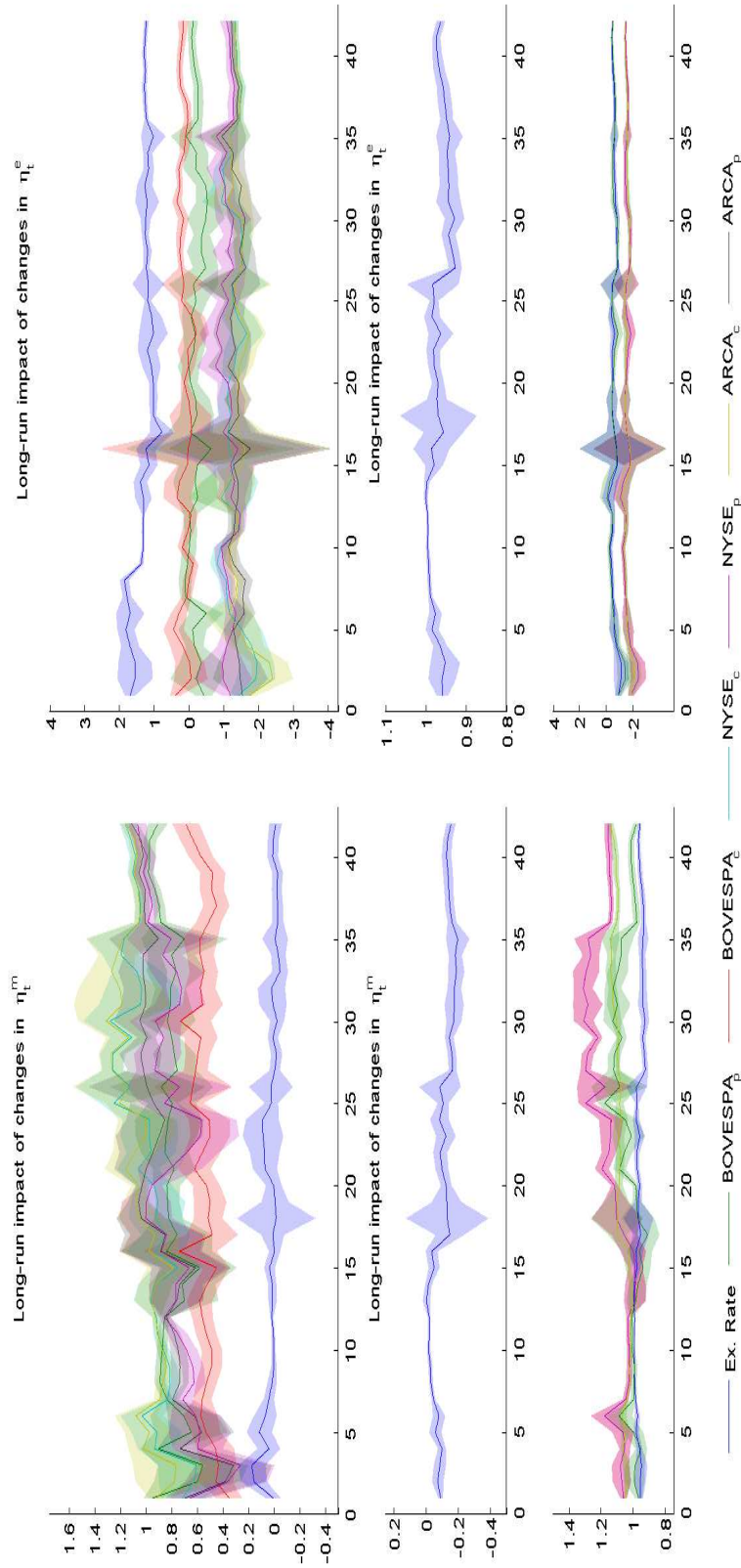
Displays rolling window estimates of the short and long-run structural polynomials, $D(0)$ and $D(1)$ respectively. Sample size and window size are fixed equal to 5,000 and 1,000 observations respectively, resulting in 39 windows. Prices are aggregated at 240 seconds. $BOVESPA_p$ accounts for preferred shares traded at BOVESPA (Brazil), $NYSE_p$ for the ADR's on preferred shares traded at NYSE and $ARCA_p$ for preferred shares traded at ARCA. Exchange rate (ExRate) is expressed in R\$ per US dollars. 95% confidence intervals are obtained using the bootstrap standard errors.

Figure 5.5
Vale



Displays rolling window estimates of the short and long-run structural polynomials, $D(0)$ and $D(1)$ respectively. Sample size and window size are fixed equal to 30,547 and 7,637 observations respectively, resulting in 42 windows. Prices are aggregated at 30 seconds. $BOVESPA_p$ and $BOVESPA_c$ accounts for preferred and common shares traded at BOVESPA (Brazil) and $NYSE_p$ and $NYSE_c$ for the ADR's on preferred and common shares traded at NYSE. Exchange rate (ExRate) is expressed in R\$ per US dollars. 95% confidence intervals are obtained using the bootstrap standard errors.

Figure 5.6
Petrobras



Displays rolling window estimates of the short and long-run structural polynomials, $D(0)$ and $D(1)$ respectively. Sample size and window size are fixed equal to 30,547 and 7,637 observations respectively, resulting in 42 windows. Prices are aggregated at 30 seconds. $BOVESPA_p$ and $BOVESPA_e$ accounts for preferred and common shares traded at BOVESPA (Brazil), $NYSE_p$ and $NYSE_e$ for the ADR's on preferred and common shares traded at NYSE and $ARCA_p$ and $ARCA_e$ for preferred and common traded at ARCA. Exchange rate (ExRate) is expressed in R\$ per US dollars. 95% confidence intervals are obtained using the bootstrap standard errors.

Chapter 6

Conclusion

In the first chapter, we introduce the research topic. We point out its importance and the motivation behind such type of study. We present a brief summary of what each chapter targets.

In the second chapter, we introduce the instrumental so far used in the literature to measure price discovery. We also link the price discovery analysis with the standard asset pricing theory, by presenting the random walk model as the efficient price. We briefly summarize some important empirical findings using the tools presented.

In the third chapter, we show the steps to deal with a 'raw' high frequency data set. We use algorithms to clean the data and we deal with the non synchronous problem. We present some basic features of our data set and institutional aspects that make this data so interesting for our research question.

In the fourth chapter, we conduct a price discovery analysis for dual-class shares that trade at different markets. In particular, we focus on the common and preferred shares of Petrobras and Vale at the BM&FBovespa

and their ADR counterparts at the NYSE. Once we account for the BRL/USD exchange rate, this leads to a system with 5 variables for Vale and 7 variables for Petrobras given that we also observe transactions at the NYSE Arca for the latter. We gauge the contribution of each share class and market by means of Hasbrouck's (1995) information share measure. Unfortunately, the standard framework does not work well for large systems because the Cholesky decomposition it employs imposes *ex-ante* restrictions on which share class and market leads the price discovery process. To circumvent such a constraint, one would have to average the IS measures across all possible permutations of the variables that integrate the system. We thus develop an alternative IS measure that rests on the eigen-decomposition of the covariance matrix of the reduced-form errors. In stark contrast to the Cholesky decomposition, the spectral-based approach is order invariant and hence corresponds to an agnostic price discovery analysis that imposes no *a priori* lead-lag pattern in the price dynamics.

Examining both common and preferred shares allows us not only to gather more information about the fundamental value of the company, but also say something about the dual-class premium. The evidence we uncover for Petrobras and Vale are compatible either with the expropriation of preferred shareholders as a class or with the majority shareholder extracting private benefits from their control rights. In both cases, we identify the Brazilian government as the main beneficiary of the dual-class premium. It detains not only Petrobras' control by holding over 55% of the voting shares, but also Vale's indirect control through a consortium of state pension funds. Note that the dual-class premium is a common factor governing the dynamics of the system only in normal times given that it becomes sta-

tionary in periods of financial distress. We also find that the foreign market is more important than the home market for the price discovery in both Petrobras and Vale. As a matter of fact, we notice that the IS estimates we obtain are by a long chalk increasing with the trade intensity of the corresponding price and hence the dominance of the NYSE. This pattern actually becomes more pronounced in the aftermath of the financial crisis, with the BM&FBovespa losing much of its importance for Petrobras and Vale in this period.

As for the exchange rate, we observe that it is the ADR prices that incorporate any shock in the efficient exchange rate. Our results also indicate that the efficient exchange rate is not exogenous to changes in the fundamental values of Petrobras and Vale. We conjecture that this is an artifact due to the omission of commodity indices in the analysis. For instance, one could include international oil prices in the Petrobras' system and the S&P industrial metals spot index in the Vale's analysis. The correlation between changes in commodity prices and the exchange rate variation is normally very high and hence we predict that augmenting the systems would help recover the expected exogeneity of the exchange rate.

In fifth chapter, we investigate the price discovery for cross listed Brazilian companies in this chapter. We are interested on measuring how fast permanent innovations are impounded by the different platforms, as well as what markets are the most important on incorporating this new information.

We present a simple market microstructure model, that guides the understanding of the empirical results. Our model allows the observed prices to depend on three different common factors: the efficient exchange rate,

the efficient asset price and efficient voting premium. Moreover, we allow the common factors to be contemporaneously correlated, yielding the necessary conditions for cross linkages among the common factors. We provide short-run and long-run solutions as function of the structural parameters, as well as price discovery dynamic measures.

We propose an alternative methodology to measure instantaneous effects of permanent shocks on prices in spirit of Yan and Zivot (2007). Our methodology does not present any issues regarding normalization of the cointegrating vectors. It is order invariant and works properly even for a large number of variables and cointegrating vectors. By using the structural framework, we are able to assess whether a permanent shock on the exchange rate changes the company's value more than the expected arbitrage adjustment. This is an interesting point which so far has not been analyzed using the price discovery framework. We find through a Monte Carlo exercise that our measure presents better performance in finite sample, when compared to all competitors.

On the empirical results, we find that the trading platform ARCA is the most efficient market on incorporating shocks instantaneously. US market is the one that adjusts for exchange rate shocks. We also observe that liquidity plays an important role on how fast markets impound new information.

Finally, the theoretical model proposed leads to interesting results regarding the price process. We find that in real terms, Brazilian companies lose value in the "long-run", following a depreciation of the Brazilian currency and a positive innovation in the efficient voting premium leads to an increase in the asset's value.

In summary this thesis presents contributions in two sides: method-

ological and empirical. In terms of methodology we create two alternative methods to measure price discovery, solving the usual ordering problem in the literature. In Chapter 5, we not only solve the ordering issue, but also deal with a structural model, where we are able to identify the sources of each innovation on prices. Regarding the empirical contribution, we adopt a novel data set, that presents a large advantage to measure price discovery when compared to other studies in the literature. Our findings are indeed interesting not only in terms of the standard price discovery analysis, but also in terms of common and preferred shares and the exchange rate role in this process.

Bibliography

- AMIHUD, Y., AND H. MENDELSON (1987): “Trading Mechanisms and Stock Returns: An Empirical Investigation,” *The Journal of Finance*, 42, 533–553.
- BAILLIE, R. T., G. G. BOOTH, Y. TSE, AND T. ZABOTINA (2002): “Price Discovery and common factor models,” *Journal of Financial Markets*, 5, 309–321.
- BARCLAY, M., AND T. HERDERSHOTT (2003): “Price Discovery and Trading after hours,” Discussion paper, University of Rochester and University of California, Berkeley.
- BARUCH, S., G. A. KAROLYI, AND M. L. LEMMON (2007): “Multimarket Trading and Liquidity: Theory and Evidence,” *The Journal of Finance*, LXII, 2169–2200.
- BOOTH, G. G., R. W. SO, AND Y. TSEH (1999): “Price Discovery in the German Equity Index Derivatives Markets,” *The Journal of Futures Markets*, 19, 619–643.
- BROWNLEES, C., AND G. GALLO (2006): “Financial econometric analysis at ultra-high frequency: Data handling concerns,” *Computational Statistics & Data Analysis*, 51, 2232–2245.

- CHU, Q. C., W. G. HSIEH, AND Y. TSE (1999): “Price discovery on the S&P 500 index markets: An analysis of spot index, index futures and SPDRs,” *International Review of Financial Analysis*, 8, 21–34.
- COCHRANE, JOHN, H. (1991): “Univariate vs. Multivariate Forecasts of GNP Growth and Stock Returns: Evidence and Implications for the Persistence of shocks, detrending methods, and tests of the permanent income hypothesis,” Discussion paper, NBER working paper, No 3427.
- DOIDGE, C., G. A. KAROLYI, AND R. M. STULZ (2001): “Why are foreign firms listed in the U.S. worth more?,” *Journal of Financial Economics*, 71, 205–238.
- FAMA, E. F. (1965): “Random Walk in Stock Market Prices,” *Financial Analysts Journal*, pp. 55–59.
- (1970): “A Review of Theory and Empirical Work,” *Journal of Finance*, 25, 383–417.
- FOERSTER, S. R., AND G. A. KAROLYI (1999): “The Effects of Market Segmentation and Investor Recognition on Asset Prices: Evidence from Foreign Stocks Listing in the United States,” *The Journal of Finance*, 54, 981–1013.
- FRIJNS, B., AND P. SCHOTMAN (2009): “Price discovery in tick time,” *Journal of Empirical Finance*, 16, 759–776.
- GAGNON, L., AND G. A. KAROLYI (2010): “Multi-market trading and arbitrage,” *Journal of Financial Economics*, 97, 5380.

- GARBADE, K. D., AND W. L. SILBER (1979): “Dominant and Satellite Markets: a study of Dually-Traded Securities,” *The Review of Economics and Statistics*, 61, No 3, 455–460.
- (1983): “Price movements and price discovery in futures and cash markets,” *Review of Economics and Statistics*, 65, 289–297.
- GONZALO, J., AND C. GRANGER (1995): “Estimation of Common Long-Memory Components in Cointegrated Systems,” *Journal of Business & Statistics*, 13, 27–35.
- GONZALO, J., AND S. NG (2001): “A systematic framework for analyzing the dynamic effects of permanent and transitory shocks,” *Journal of Economics Dynamics & Control*, 25, 1527–1546.
- GRAMMIG, J., M. MELVIN, AND C. SCHLAG (2005): “Internationally cross-listed stock prices during overlapping trading hours: price discovery and exchange rate effects,” *Journal of Empirical Finance*, 12, 139–164.
- GRAMMIG, J., AND F. J. PETER (2012): “Tell-tale tails: A data-driven approach to estimate unique market information shares,” Discussion paper, forthcoming in the *Journal of Financial and Quantitative Analysis*.
- HAMILTON, J. D. (1994): *Time Series Analysis*. Princeton University Press.
- HARRIS, F. H., T. H. MCINISH, G. L. SHOESMITH, AND R. A. WOOD (1995): “Cointegration, Error Correction, and Price Discovery on Informationally Linked Security Markets,” *Journal of Financial and Quantitative Analysis*, 4, 563–578.

- HARRIS, F. H., T. H. MCINISH, AND R. A. WOOD (2002): “Security price adjustment across exchanges: an investigation of common factor components for Dow stocks,” *Journal of Financial Markets*, 5, 277–308.
- HASBROUCK, J. (1993): “Assessing the Quality of a Security Market: A New Approach to Transaction-Cost Measurement,” *Review of Financial Studies*, 6, 191212.
- HASBROUCK, J. (1995): “One Security, Many Markets: Determining the Contributions to Price Discovery,” *The Journal of Finance*, 50, 1175–1198.
- (2002): “Stalking the efficient price in market microstructure specifications: an overview,” *Journal of Financial Markets*, 5, 329–339.
- (2003): “Intraday Price Formation in U.S. Equity Index Markets,” *The Journal of Finance*, 58, 2375–2399.
- (2007): *Empirical Market Microstructure: The Institutions, Economics, and Econometrics of Securities Trading*. Oxford University Press.
- HASBROUCK, J., AND T. HO (1987): “Order Arrival, Quote Behavior, and the Return-Generating Process,” *The Journal of Finance*, 42, 1035–1048.
- HUANG, R. D. (2002): “The quality of ECN and Nasdaq market maker quotes,” *Journal of Finance*, 57, 1285–1319.
- HUPPERETS, E. C., AND A. J. MENKVELD (2002): “Intraday analysis of market integration: Dutch blue chips traded in Amsterdam and New York,” *Journal of Financial Markets*, 53, 57–82.

- JOHANSEN, S. (1988): “Statistical Analysis of Cointegration Vectors,” *Journal of Economic Dynamics and Control*, 12, 231–254.
- (1991): “Estimation and Hypothesis Testing of Cointegration Vectors in Gaussian Vector Autoregressive Models,” *Econometrica*, 59, 1551–1580.
- JONG, D. F., AND P. C. SCHOTMAN (2010): “Price Discovery in Fragmented Markets,” *Journal of Financial Econometrics*, 8, 1–28.
- JONG, F. D. (2002): “Measures of contributions to price discovery: A comparison,” *Journal of Financial Markets*, 5, 323–327.
- JONG, DE F., M. R. S. P. (1998): “Price discovery in the foreign exchange market: an empirical analysis of the yen/dmark rate,” *Journal of International Money and Finance*, 17, 5–28.
- KAROLYI, G. A. (2004): “The Role of ADRs in the Development of Emerging Equity Markets,” *Review of Economics and Statistics*, 86, 670–690.
- KIM, C. (2010a): “Cross-Border Listing and Price Discovery: Canadian Blue Chips Traded in TSX and NYSE,” Discussion paper, University of Washington.
- KIM, C. H. (2010b): “Price Discovery in Crude Oil Prices,” Discussion paper, University of Washington.
- LEHMANN, B. N. (2002): “Some desiderata for the measurement of price discovery across markets,” *Journal of Financial Markets*, 5, 259–276.
- LI, H., AND G. S. MADDALA (1997): “Bootstrapping cointegrating regressions,” *Journal of Econometrics*, 80, 297–318.

- LIEN, D., AND K. SHRESTHA (2009): “A new information share measure,” *Journal of Futures Markets*, 4, 377–395.
- LUTKEPOHL, H. (2007): *New Introduction to Multiple Time Series Analysis*. Springer.
- MARTENS, M. (1998): “Price discovery in high and low volatility periods: open outcry versus electronic trading,” *Journal of International Financial Markets, Institutions and Money*, 8, 243–260.
- MEI, J., J. A. SCHEINKMAN, AND W. XIONG (2005): “Speculative Trading and Stock Prices: Evidence from Chinese A-B Share Premia,” Discussion paper, National Bureau of Economic Research.
- MULLER, A., AND W. VERSCHOOR (2006): “Foreign exchange rate exposure: Survey and suggestions,” *Journal of Multinational Financial Management*, 16, 385–410.
- ROLL, R. (1984): “A Simple Implicit Measure of the Effective Bid-Ask Spread in an Efficient Market,” *The Journal of Finance*, 39, 1127–1139.
- SMITH, B. F., AND B. AMOAKO-ADU (1995): “Relative prices of dual class shares,” *Journal of Financial and Quantitative Analysis*, 30, 563–578.
- STOCK, J. H., AND M. W. WATSON (1988): “Testing for Common Trends,” *Journal of the American Statistical Association*, 83, No 404, 1097–1107.
- THEISSEN, E. (2005): “Price Discovery in floor and screen trading systems,” *Journal of Empirical Finance*, 9, 455–474.

- TSE, Y., P. BANDYOPADHYAY, AND Y.-P. SHEN (2006): “Intraday Price Discovery in the DJIA Index Markets,” *Journal of Business Finance & Accounting*, 33(9), 1572–1585.
- WARNE, A. (1993): “A common trends model: identification, estimation and inference,” Discussion paper, Stockholm School of Economics.
- YAN, B., AND E. ZIVOT (2007): “The Dynamics of Price Discovery,” Discussion paper, Department of Economics, University of Washington, Seattle, WA.
- YAN, B., AND E. ZIVOT (2010): “A structural analysis of price discovery measures,” *Journal of Financial Markets*, 13, 1–19.
- ZINGALES, L. (1994): “The value of the voting right: A study of the Milan Stock Exchange experience,” *The Review of Financial Studies*, 7, 125–148.
- (1995): “What determines the value of corporate votes?,” *The Quarterly Journal of Economics*, Nov, 1047–1073.